

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 21:20:15 ; Search time 6482 Seconds

(without alignments)  
10939.402 Million cell updates/sec

Title: US-10-018-604-1

Perfect score: 1636

Sequence: 1 aactttttcaatagacaag.....atatgaaaaaaaaaaaaa 1636

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rtd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vit.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	1636	6 A24194	A24194 L.esculentu
2	1636	100.0	1636	6 AR364905	AR364905 L.esculentu
3	1621	99.1	1621	6 A15981	A15981 L.esculentu
4	1621	99.1	1621	6 AX062336	AX062336 Sequence
5	1621	99.1	1621	6 AX062336	AX062336 Tomato mRNA
6	1614	98.6	1614	6 LEPG2AR	X05656 Tomato mRNA
7	1613.2	98.6	1624	6 LEPGR	101809 Sequence 1
8	599	34.8	599	8 CPA505750	AJ505750 Carica pa
9	570	34.6	602	8 LES505947	AJ505947 Lycopersi
10	549.2	33.6	1541	8 AF152758	AF152758 Actinidia
11	525.2	32.1	1595	8 AY043233	AY043233 Vitis vin
12	462	28.2	7456	8 TOMPGAAA	M37304 Tomato poly
13	462	28.2	7456	8 LEPOLYGA	X14074 Tomato gene
14	449.2	27.5	1600	8 AK117942	AK117942 Arabidops
15	445.2	27.2	1296	6 AX412563	AX412563 Sequence
16	445.2	27.2	1296	6 AX651724	AX651724 Sequence
17	445.2	27.2	1296	6 BT005376	BT005376 Arabidops
18	439.8	26.9	1725	8 AVOPOLYCAL	L06094 Avocado pol
19	433	26.5	1725	8 PAPOLYGV	X66426 P.americana
20	431.4	26.4	1364	8 AY078936	AY078936 Arabidops
21	429.6	26.3	1765	8 BNA250919	AB250919 Brassica
22	428.2	26.2	1767	8 AF062467	AF062467 Cucumis m
23	420.4	25.7	1657	6 A57101	A57101 Sequence 5
24	420.4	25.7	1657	6 AX016328	AX016328 Sequence
25	420.4	25.7	1657	6 BNSAC66	Z49971 B.napus of
26	420.2	25.7	1599	8 BNPGLACR	X95800 B.napus mRN
27	420.2	25.7	1631	6 A62402	A62402 Sequence 1
28	414.6	25.3	1302	8 BRA428543	A428543 Brassica
29	405.8	24.3	1208	8 MAUP14A	L27743 Malus domes
30	397.6	24.3	1208	8 RIATJ4147	AJ224147 Rubus ida
31	396.2	24.2	1795	8 AB084461	AB084461 Pyrus com
32	395	24.1	414	8 TOMPGA	M02069 L.esculentu
33	391.6	23.9	1280	6 AX651722	AX651722 Sequence
34	375	22.9	1269	6 AX653992	AX653992 Sequence
35	369	22.6	1182	6 AX653993	AX653993 Sequence
36	331.8	20.3	2207	6 I09042	I09042 Sequence 5
37	331.8	20.3	2207	8 LEPG5	X07410 Lycopersico
38	308.4	18.9	1603	8 AF138858	AF138858 Lycopersi
39	296	18.1	1344	6 AX654086	AX654086 Sequence
40	278.8	17.0	7471	8 FCCCHRAF	L47242 Capsicum an
41	275.8	16.9	573	8 AF152752	AF152752 Actinidia
42	275.8	16.9	667	8 AF152753	AF152753 Actinidia
43	275.2	16.8	690	8 AF152755	AF152755 Actinidia
44	275.2	16.8	696	8 AF152754	AF152754 Actinidia
45	272	16.6	1366	8 AY142668	AY142668 Arabidops

# ALIGNMENTS

RESULT 1

A24194 1636 bp DNA linear PAT 04-APR-1995

LOCUS A24194

DEFINITION L.esculentum polygalacturonase clone ptom6.

ACCESSION A24194

VERSION A24194.1 GI:904404

# KEYWORDS

# SOURCE

Lycopersicon esculentum (tomato)

# ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

# REFERENCE

1 (bases 1 to 1636)  
Bridges, L.G., Grierson, D. and Schuch, W.W.

# AUTHORS

Anti-sense regulation of plant gene expression

JOURNAL Patent: EP 0532060-A 1 17-MAR-1993:

IMPERIAL CHEMICAL INDUSTRIES PLC; ZENECA LIMITED

FEATURES

source

Location/Qualifiers

1..1636

/organism="Lycopersicon esculentum"

/mol\_type="unassigned DNA"

/db\_xref="taxon:4081"

51..1424

CDS

/codon\_start=1

/product="polygalacturonase"

/protein\_id="CAA01720.1"

/db\_xref="GI:904405"

/db\_xref="GOA:P05117"

/db\_xref="SWISS-PROT:P05117"

/translation="VVIQNSILLIITFASISICRSNVIDNLFKQYDNLICEF

AHDFOALSYLSKINENNNIDKDKNGKIVNLVSPGKGGKTYDNLAEQAWNEA

CSSRTVQFVPEKKNRYLLKQTFESGPKSSISVKI FGLSESSKISDYKRLMIAF

DSVQMLVVGGGGTINGNQWVWSSCKINSLPCRDAPALFPMCKMLKVNLSKN

AQOIHKESTNVASNMIMASAKSPRTDGVHVNQYIQISDITIGTGDKLSIV

SGSONVOAINTICGPHGISIGLSGNSSEAVSVTVNEAKIIGAENGVRITKMOGG

SGAENIKELNEMODVYKPIIIDQNYCDRVPCTCOQPSAQQVKNVVENIKETSAK

VAIKRDCSTNPFCEGIIMENINLVESGKPSBATCKVHFNNAEHVTPHCTSLISED

EALVINY"

ORIGIN

Query Match 100.0% Score 1636; DB 6; Length 1636;

Best Local Similarity 100.0% Pred. No. 2.9e-246;

Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AATCTTTTCATAGCAAGTTTAAACCATACATATACATATATATGCTTATCC 60  
1 AATCTTTTCATAGCAAGTTTAAACCATACATATACATATATATGCTTATCC 60  
61 AAAGGATAGTATCTCTCTCTCATATTTATTTGCTTCATCATTTCACTTGTAGA 120  
61 AAAGGATAGTATCTCTCTCTCATATTTATTTGCTTCATCATTTCACTTGTAGA 120  
121 GCATGTTATGATGACAAATTTATTCACCAAGTTTATGATTAATTTCTTGAACAAGAT 180  
121 GCATGTTATGATGACAAATTTATTCACCAAGTTTATGATTAATTTCTTGAACAAGAT 180  
181 TTGTCATGATTTTCAAGCTTATCTTCTTATTTGACCAAAATATTTGAACAACAATA 240  
181 TTGTCATGATTTTCAAGCTTATCTTCTTATTTGACCAAAATATTTGAACAACAATA 240  
241 AATTTGACAAAGTTGATTAATAATGGGATTAAGTATTAATGACTTACCTTGAAGCTA 300  
241 AATTTGACAAAGTTGATTAATAATGGGATTAAGTATTAATGACTTACCTTGAAGCTA 300  
301 AGGATGATGACAAACATATGATTAATTTGATTTGACCAAGCAATGGAATGAGATGT 360  
301 AGGATGATGACAAACATATGATTAATTTGATTTGACCAAGCAATGGAATGAGATGT 360  
361 CATCTAGAACACCTGTTCAATTTGTTCTTAAACCAAGATTAATCTTCAAGCAAA 420  
361 CATCTAGAACACCTGTTCAATTTGTTCTTAAACCAAGATTAATCTTCAAGCAAA 420  
421 TCACCTTTTCAAGTCCATGATTTCTTCAATTTGATTAAGATTTTGTGATCTTGAAG 480  
421 TCACCTTTTCAAGTCCATGATTTCTTCAATTTGATTAAGATTTTGTGATCTTGAAG 480  
481 CATCTAGTAAATTTCAAGCTACAAAGATAGAAAGCTTTGATGCTTTTGAATGCTTTC 540  
481 CATCTAGTAAATTTCAAGCTACAAAGATAGAAAGCTTTGATGCTTTTGAATGCTTTC 540  
541 AAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGAAGCAAGATGATGTTGAGCA 600  
541 AAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGAAGCAAGATGATGTTGAGCA 600  
601 GTTCTTGGAAAAATAATTAATCACTGCAATGAGGATGACCAACGCGCTTAACTTCT 660  
601 GTTCTTGGAAAAATAATTAATCACTGCAATGAGGATGACCAACGCGCTTAACTTCT 660

661 GGAATTTGCAAAATTTGAAAGTAAATATCTAAAGGTAATAATGACACAAATTCATA 720  
661 GGAATTTGCAAAATTTGAAAGTAAATATCTAAAGGTAATAATGACACAAATTCATA 720  
721 TCAAAATTTGAGTCAATGCACTAATGTTGAGCTTCAAAATTTGATGATCAATGCTTCA 780  
721 TCAAAATTTGAGTCAATGCACTAATGTTGAGCTTCAAAATTTGATGATCAATGCTTCA 780  
781 AGAGCCCAATATCTGATGAGTCCATGATTAATTAATTAATTAATTAATTAATTAAT 840  
781 AGAGCCCAATATCTGATGAGTCCATGATTAATTAATTAATTAATTAATTAATTAAT 840  
841 AGAGCCCAATATCTGATGAGTCCATGATTAATTAATTAATTAATTAATTAATTAAT 840  
841 AGAGCCCAATATCTGATGAGTCCATGATTAATTAATTAATTAATTAATTAATTAAT 840  
841 CATATTATGGAACAGGTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 900  
841 CATATTATGGAACAGGTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 900  
901 CCACAAATTAATTAATTTGATGATGATGATTAATTAATTAATTAATTAATTAATTAAT 960  
901 CCACAAATTAATTAATTTGATGATGATGATTAATTAATTAATTAATTAATTAATTAAT 960  
961 ATTGCAAGCTTATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
961 ATTGCAAGCTTATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
1021 ATGAGTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080  
1021 ATGAGTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080  
1081 TGAATGTTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140  
1081 TGAATGTTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140  
1141 GAGTTGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200  
1141 GAGTTGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200  
1201 GAGTTGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
1201 GAGTTGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
1261 CATGTTGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
1261 CATGTTGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
1321 AGGCTACGTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1321 AGGCTACGTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1381 AGGCTACGTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440  
1381 AGGCTACGTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440  
1441 CATGTTGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
1441 CATGTTGGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
1501 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
1501 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
1561 GTCAAAAGTTGACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620  
1561 GTCAAAAGTTGACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620  
1621 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680  
1621 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680

RESULT 2  
AR364905  
LOCUS  
DEFINITION Sequence 2 from patent US 5447867.

[illegible][illegible]

AUTHORS Bridges, I.G., Schuch, W.W. and Grierson, D.  
 TITLE Anti-sense regulation of plant gene expression  
 JOURNAL Patent: EP 0271988-A 1 22-JUN-1988;  
 IMPERIAL CHEMICAL INDUSTRIES PLC

## FEATURES

Location/Qualifiers  
 1..1621

/organism="Lycopersicon esculentum"  
 /mol\_type="unassigned RNA"  
 /strain="Ailsa Craig"  
 /db\_xref="taxon:4081"  
 /clone="PTOM 6"  
 /issue\_type="fruit"  
 /dev\_stage="fruit ripe"  
 51..1424  
 /codon\_start=1  
 /product="polygalacturonase-2a"  
 /protein\_id="CA01256.1"  
 /db\_xref="GI:490040"  
 /db\_xref="GOA:P05117"  
 /db\_xref="SWISS-PROT:P05117"

## CDS

/translation="MYQNSILILILIFASSISTCRSNVIDDNLFKQYVDNLDEOF  
 AHDFQAYLSYLSKNIESNNIDKVDKGIKVINLSFGKGGKTYDNIAPQAMREA  
 CSSRTPVQFVVRKKNRYLIKQTTPSPCRSSISVKI FGLSLSKSI SDYKDRMTAF  
 DSVQNLVGGGGTINSGVWMPSSCKLNKSLPCRDAPALTFMCKLNKNNLSKN  
 AQQIH KESCTNVVASNLMINASAKSPMTDGHVSNITQYIISDIIIGTDICISIV  
 SSGSNVOATNITCGPHGISIGLSGNSNAVYVTEAKIIGLENVRITKTGG  
 SGOASNKILNEMODVKYPIILIDNYCDRVEPCIOQFSAVOVKVVENIGTSATX  
 VALKDCSTNFPCEGIMENITLVSGSRSEATCKNHFNNAEHVTHICTSLSESD  
 EALLNYN"

mat\_peptide 264..1421  
 /product="polygalacturonase-2a"  
 misc\_feature 615..623  
 /note="pot. N-glycosylation site"  
 misc\_feature 768..776  
 /note="pot. N-glycosylation site"  
 misc\_feature 906..914  
 /note="pot. N-glycosylation site"  
 misc\_feature 981..989  
 /note="pot. N-glycosylation site"  
 misc\_feature 1598..1606  
 /note="pot. polyA signal"  
 polyA\_site 1621  
 /note="polyA site"

## ORIGIN

Query Match 99.1%; Score 1621; DB 6; Length 1621;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-244;  
 Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AATCTTTTCATATGACAAAGTTTAAAAACCATACATATACATATATATGTTATCC 60  
 1 AATCTTTTCATATGACAAAGTTTAAAAACCATACATATATATATATGTTATCC 60  
 61 AAAGAAATGATATCTCTCTTCATATATATTTTGGTTCATCAATTTGACATTTGAGA 120  
 61 AAAGAAATGATATCTCTCTTCATATATATTTTGGTTCATCAATTTGACATTTGAGA 120  
 121 GCAATGTTATGATGACATATATTTCAACAGTTTATGATATATCTTGAACAAGAT 180  
 121 GCAATGTTATGATGACATATATTTCAACAGTTTATGATATATCTTGAACAAGAT 180  
 181 TTGCTCATGATTTCAAGTTTATCTTTCTTATTTGAGCAAAAATATTTGAACAACAATA 240  
 181 TTGCTCATGATTTCAAGTTTATCTTTCTTATTTGAGCAAAAATATTTGAACAACAATA 240  
 241 ATATTGACAAAGTTTATATAAAATGGAATTAAGTATTAATGACTTTGAGAGCTA 300  
 241 ATATTGACAAAGTTTATATAAAATGGAATTAAGTATTAATGACTTTGAGAGCTA 300  
 301 AGGATGATGAAAAACATATGATATATTTGATTTGAGCAAGCATGAAATGAAGATGTT 360  
 301 AGGATGATGAAAAACATATGATATATTTGATTTGAGCAAGCATGAAATGAAGATGTT 360

361 CATCTAGAAACACCTGTTCAATTTGGTTCCTTAAAAACAAGATTAATCTTCAAGCAAA 420  
 361 CATCTAGAAACACCTGTTCAATTTGGTTCCTTAAAAACAAGATTAATCTTCAAGCAAA 420  
 421 TCACCTTTTCAGGTCATGACAGATCTTCAATTTGAGTAAAGATTTTGGATCCCTAGAG 480  
 421 TCACCTTTTCAGGTCATGACAGATCTTCAATTTGAGTAAAGATTTTGGATCCCTAGAG 480  
 481 CATCTAGTAAATTTTACGACTACAAAGATTAAGAGGCTTTGATGCTTTGATAGTTC 540  
 481 CATCTAGTAAATTTTACGACTACAAAGATTAAGAGGCTTTGATGCTTTGATAGTTC 540  
 541 AAAATTTAGTTGTGAGAGAGAGAACTATCAATGGCATGAGCAAGTATGGTGGCCAA 600  
 541 AAAATTTAGTTGTGAGAGAGAGAACTATCAATGGCATGAGCAAGTATGGTGGCCAA 600  
 601 GTTCTTGCAAAATTAATAATCACTGCCATGACAGGATGACACCAAGGCTTAACCTTCT 660  
 601 GTTCTTGCAAAATTAATAATCACTGCCATGACAGGATGACACCAAGGCTTAACCTTCT 660  
 661 GGAATTTGCAAAATTTGAAAGGAAATTAATCTAAAGTAAAAATGACAACAAAATTCATA 720  
 661 GGAATTTGCAAAATTTGAAAGGAAATTAATCTAAAGTAAAAATGACAACAAAATTCATA 720  
 721 TCAAAATTTGAGTCATGACACTAATGTTGTAGCTTCAAAATTTGATGATCAATGCTTCAGAA 780  
 721 TCAAAATTTGAGTCATGACACTAATGTTGTAGCTTCAAAATTTGATGATCAATGCTTCAGAA 780  
 781 AAGAGCCCAATATCTGATGAGTCCATGTAATCTCAATATATCTCAATATATCTGATA 840  
 781 AAGAGCCCAATATCTGATGAGTCCATGTAATCTCAATATATCTCAATATATCTGATA 840  
 841 CTATTTATTTGGAACAGGTATGATGATGATTTCAATGTTCTGATCTCAAAATTTGACAAAG 900  
 841 CTATTTATTTGGAACAGGTATGATGATGATTTCAATGTTCTGATCTCAAAATTTGACAAAG 900  
 901 CCACAAATATTAATCTGTTGTCAGGTCATGATGATTAAGTATGGAAGCTTAGATCTGAAA 960  
 901 CCACAAATATTAATCTGTTGTCAGGTCATGATGATTAAGTATGGAAGCTTAGATCTGAAA 960  
 961 ATTCAAGAGCTTATGCTGTAATGTTACTGTAATGAAAGCCAAATTTATGCTGCGGAAA 1020  
 961 ATTCAAGAGCTTATGCTGTAATGTTACTGTAATGAAAGCCAAATTTATGCTGCGGAAA 1020  
 1021 ATGAGAGTTAGATTCAGAGCTTGGCAGGAGATCTGGAACAACCTTGCAATCAATTTTC 1080  
 1021 ATGAGAGTTAGATTCAGAGCTTGGCAGGAGATCTGGAACAACCTTGCAATCAATTTTC 1080  
 1081 TGAATGTGGAATGCAAGCGTTAAGTATCCATATTAATAGACCAAACTAATTTGATC 1140  
 1081 TGAATGTGGAATGCAAGCGTTAAGTATCCATATTAATAGACCAAACTAATTTGATC 1140  
 1141 GAGTTGAACCATGTAATACAAAGTTTTCAGAGTTCAAGTGAATAATGAGTATGAGAGA 1200  
 1141 GAGTTGAACCATGTAATACAAAGTTTTCAGAGTTCAAGTGAATAATGAGTATGAGAGA 1200  
 1201 ATATCAAGGGGCAAGGTGCAACAAAGGTGGCAATAAATTTGATGACACAAATTTTC 1260  
 1201 ATATCAAGGGGCAAGGTGCAACAAAGGTGGCAATAAATTTGATGACACAAATTTTC 1260  
 1261 CATGTGAAGAAATTAATGAGAAATTAATTTAGTGGGAAAGTGGAAAAACCATCAG 1320  
 1261 CATGTGAAGAAATTAATGAGAAATTAATTTAGTGGGAAAGTGGAAAAACCATCAG 1320  
 1321 AGGCTACGTGCAAAAATGCTCTTTTAACATGCTGAACATGTTACACCACTGACCTT 1380  
 1321 AGGCTACGTGCAAAAATGCTCTTTTAACATGCTGAACATGTTACACCACTGACCTT 1380  
 1381 CACTGAAATTTCAAGAGTGAAGCTCTTTGTATATTAATTTATATCTATGATGATCT 1440  
 1381 CACTGAAATTTCAAGAGTGAAGCTCTTTGTATATTAATTTATATCTATGATGATCT 1440  
 1441 TCAATATATGACAGATGATATATATCAAAATTAATCTATGATGATGATGATGAT 1500



Db 1441 TCATATATATGCGATATGATATATCATATTAACAATCTATATCATATGATTAATGAATAA 1500

Qy 1501 TTATTTATTAATATGATGATGATGATTAATTAATAGACTATGATTTCTTATTTCTTA 1560

Db 1501 TTATTTATTAATATGATGATGATGATTAATTAATAGACTATGATTTCTTATTTCTTA 1560

Qy 1561 GTCAAAAGTTTGAAGATTTGATCTTTTATGATGACAAAATATATTAATGCTTATTTATAT 1620

Db 1561 GTCAAAAGTTTGAAGATTTGATCTTTTATGATGACAAAATATATTAATGCTTATTTATAT 1620

Qy 1621 G 1621

Db 1621 G 1621

RESULT 4  
AX062336  
LOCUS AX062336 1621 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 1 from Patent WO0078982.  
ACCESSION AX062336  
VERSION AX062336.1 GI:12540229  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Christensen, T.M. and Kreiberg, J.D.  
TITLE Process for the enzymatic modification of pectin  
JOURNAL Patent: WO 0078982-A 1 28-Dec-2000;  
DANISCO A/S (DK)  
FEATURES  
source location/Qualifiers  
1. 1621  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="cDNA clone"

ORIGIN  
Query Match 99.1%; Score 1621; DB 6; Length 1621;  
Best Local Similarity 100.0%; Pred. No. 6,4e-244; Indels 0; Gaps 0;  
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AATCTTTTTCATATGACAGATTTTAAAAACATACCATATTAACATATATCATGCTTATCC 60  
Db 1 AATCTTTTTCATATGACAGATTTTAAAAACATACCATATTAACATATATCATGCTTATCC 60

Qy 61 AAAGGAATGATATCTCTCTCTCATATATTTTGGCTTCATCATTTCAACTGTGAGA 120  
Db 61 AAAGGAATGATATCTCTCTCTCATATATTTTGGCTTCATCATTTCAACTGTGAGA 120

Qy 121 GCATGTTATGATGACATTTTCAACAAGTTTATGATTAATTTGTTGACAAGAT 180  
Db 121 GCATGTTATGATGACATTTTCAACAAGTTTATGATTAATTTGTTGACAAGAT 180

Qy 181 TTGCTCATGATTTCAAGCTTATCTTCTTATTTGAGCAAAAATATGAAAGCAACATA 240  
Db 181 TTGCTCATGATTTCAAGCTTATCTTCTTATTTGAGCAAAAATATGAAAGCAACATA 240

Qy 241 ATATTTGACAGGTTGATTAATAATGGGATTAAGTGATTAATGATCTTAAGCTTTGAGCTA 300  
Db 241 ATATTTGACAGGTTGATTAATAATGGGATTAAGTGATTAATGATCTTAAGCTTTGAGCTA 300

Qy 301 AGGTTGATGAAAAAATATATATATGATTTGAGCAAGCAAGCAATGAGATGAGATGTT 360  
Db 301 AGGTTGATGAAAAAATATATATATGATTTGAGCAAGCAAGCAATGAGATGAGATGTT 360

Qy 361 CATCTAGAACACCTGTTCAATTTTGTGTTCTTAAAAACAAGATTAATCTTCAAGCAAA 420  
Db 361 CATCTAGAACACCTGTTCAATTTTGTGTTCTTAAAAACAAGATTAATCTTCAAGCAAA 420

Qy 421 TCACCTTTTCAGGTCATGACGATTTTCTATTTTCATTAAGATTTTGGATCTTAAAG 480  
Db 421 TCACCTTTTCAGGTCATGACGATTTTCTATTTTCATTAAGATTTTGGATCTTAAAG 480

Db 421 TCACCTTTTCAGGTCATGACGATTTTCTATTTTCATTAAGATTTTGGATCTTAAAG 480

Qy 481 CATCTAGAACAAATTTTCAAGCTTAAAGATTAAGAGCTTTGATTTGCTTTGATATGCTTC 540

Db 481 CATCTAGAACAAATTTTCAAGCTTAAAGATTAAGAGCTTTGATTTGCTTTGATATGCTTC 540

Qy 541 AAAATTTAGTTTGGAGAGAGAGAACTTCAATGGCAATGACAAAGTATGGGGGCA 600

Db 541 AAAATTTAGTTTGGAGAGAGAGAACTTCAATGGCAATGACAAAGTATGGGGGCA 600

Qy 601 GTTCTTGCAAAATTAATTAATCACTGCAAGGAGATGACCAACGAGCTTAACTTCT 660

Db 601 GTTCTTGCAAAATTAATTAATCACTGCAAGGAGATGACCAACGAGCTTAACTTCT 660

Qy 661 GGAATTCGCAAAATTTGAAAGTGAATATCTAAAGATTAAGATTAAGCAACAAATCTATA 720

Db 661 GGAATTCGCAAAATTTGAAAGTGAATATCTAAAGATTAAGATTAAGCAACAAATCTATA 720

Qy 721 TCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780

Db 721 TCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780

Qy 781 AGAGCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

Db 781 AGAGCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

Qy 841 CTATTTATTTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

Db 841 CTATTTATTTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

Qy 901 CCACCAATATTTACTGTTGGTCCAGGTATGATGATGATGATGATGATGATGATGATGAT 960

Db 901 CCACCAATATTTACTGTTGGTCCAGGTATGATGATGATGATGATGATGATGATGATGAT 960

Qy 961 ATTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020

Db 961 ATTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020

Qy 1021 ATGAGATTTAGGATCAAGATCTTGGCAGGAGATCTGACAGCTAGCAACATCAAAATTTTC 1080

Db 1021 ATGAGATTTAGGATCAAGATCTTGGCAGGAGATCTGACAGCTAGCAACATCAAAATTTTC 1080

Qy 1081 TGAATTTGGAATTCAGAGCTTAAAGTATGATGATGATGATGATGATGATGATGATGAT 1140

Db 1081 TGAATTTGGAATTCAGAGCTTAAAGTATGATGATGATGATGATGATGATGATGATGAT 1140

Qy 1141 GAGTTGAACATGATTAACAAGCTTTTCAAGAGTTTCAAGTGAATTAATGATGATGATGAT 1200

Db 1141 GAGTTGAACATGATTAACAAGCTTTTCAAGAGTTTCAAGTGAATTAATGATGATGATGAT 1200

Qy 1201 ATATCAAGGAGCAAGTGCACAAAGGTGGCATTAATTTGATTTGACGACAAACCTTTC 1260

Db 1201 ATATCAAGGAGCAAGTGCACAAAGGTGGCATTAATTTGATTTGACGACAAACCTTTC 1260

Qy 1261 CATGTGAAGCAATTAATGAGAAATTAATTTAGTGGGAAAGTGGAAACCAATCAG 1320

Db 1261 CATGTGAAGCAATTAATGAGAAATTAATTTAGTGGGAAAGTGGAAACCAATCAG 1320

Qy 1321 AGGCTAGGTGCAAAAATGTCATTTTAAATGCTGAATGCTTAACTTAACTTAACTTAACTT 1380

Db 1321 AGGCTAGGTGCAAAAATGTCATTTTAAATGCTGAATGCTTAACTTAACTTAACTTAACTT 1380

Qy 1381 CACTAGAAATTTCAAGAGATGAGCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 1440

Db 1381 CACTAGAAATTTCAAGAGATGAGCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 1440

Qy 1441 TCATATTTATGACATATGAT 1500

Db 1441 TCATATTTATGACATATGAT 1500

Qy 1501 TTATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560

Db 1501 TTATTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560

QY 1561 GTCAAAAGTTGACGATTTTAAATGACAAAATAAATAAGTTATTAAT 1620  
Db 1561 GTCAAAAGTTGACGATTTTAAATGACAAAATAAATAAGTTATTAAT 1620  
QY 1621 G 1621  
Db 1621 G 1621

RESULT 5  
LEPG2AR 1621 bp mRNA linear PLN 27-MAR-1995  
LOCUS Tomato mRNA for polygalacturonase-2a (PG-2a).  
DEFINITION X04583  
ACCESSION X04583.1 GI:19291  
VERSION glycoprotein; polygalacturonase; polygalacturonase-2a; signal  
KEYWORDS peptide;  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 1621)  
Grierson, D., Tucker, G.A., Keen, J., Ray, J., Bird, C.R. and Schuch, W.  
Sequencing and identification of a cDNA clone for tomato  
polygalacturonase  
Nucleic Acids Res. 14 (21), 8595-8603 (1986)  
87066731  
3786135

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

FEATURES  
source  
1..1621  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/strain="Alisa Craig"  
/db\_xref="taxon:4081"  
/clone="pTOW 6"  
/tissue\_type="fruit"  
/dev\_stage="fruit ripe"  
51..1424  
/note="unnamed protein product; precursor polypeptide (AA  
-71 to 386)"  
/codon\_start=1  
/protein\_id="CAA28254.1"  
/db\_xref="GI:19292"  
/db\_xref="GOA:P05117"  
/db\_xref="SWISS-PROT:P05117"  
/translation="VIGRNSILILITIPASSISTCRSNVIDNLFKQYDNILIEQF  
AHDFQAYLSYLSKINIESNNIDKDKNGIKVINVLSPGAGDKGTVDNIAEQMNEA  
CSSRPVQVFNKKNYLLKQITFSGPGRSSIVKIPSLASSKISDYKORBLIAF  
DSQNLVVGSGGTINGGOWPSSCKINKSLPCDAPDTALTPEWCKLTKNNLSKY  
AQQHIRESCNTVVASNLMTNASKSPGTGVAHNSQYICISITFIIGTGDICISV  
SGSQNVQATNTTCGPGHGISIGLSGNSSEAVSVNTVEAKITGAENGVIKTMGG  
SGQSNIKFLVEMQDVKPIITIDQNYCDREYPCIQPSAVQKNVVENKIGSATK  
VAIRFDCSTNPEGIIIMENINLVGSEKPEATCKNVHFNNAEHVTCHSLEISD  
EALILNY"  
264..1421  
/product="mature polygalacturonase (AA 1-386)"  
51..263  
/note="put. pre-peptide (AA -71 to -1)"  
615..623  
/note="pot. N-glycosylation site"  
768..776  
/note="pot. N-glycosylation site"  
906..914  
/note="pot. N-glycosylation site"  
misc\_feature  
981..989

misc\_feature 1598..1606 /note="pot. N-glycosylation site"  
polyA\_site /note="pot. polyA signal"  
1621  
/note="polyA site"

ORIGIN  
Query Match 99.1%; Score 1621; DB 8; Length 1621;  
Best Local Similarity 100.0%; Pred.No. 6,4e-244;  
Matches 1621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTTTCAATAGACAAAGTTTAAAAACCATATACATATATATATATATCC 60  
Db 1 AATCTTTCAATAGACAAAGTTTAAAAACCATATATATATATATATATATCC 60  
QY 61 AAAGGATGATATCTCTCTTCATATATATTTTCTTCATCAATTCACCTTGAGAA 120  
Db 61 AAAGGATGATATCTCTCTTCATATATATTTTCTTCATCAATTCACCTTGAGAA 120  
QY 121 GCAATGTTATGATGACAAATTTTCAACAGTTATATATATATATATATATAT 180  
Db 121 GCAATGTTATGATGACAAATTTTCAACAGTTATATATATATATATATATAT 180  
QY 181 TTGCTCATGATTTTCAAGCTTATCTTTCTTATTTGACAAAATATGAAAGCAATA 240  
Db 181 TTGCTCATGATTTTCAAGCTTATCTTTCTTATTTGACAAAATATGAAAGCAATA 240  
QY 241 ATATTGACAAAGTTGATTAATAATGAGATTAAATGATATATGATCTTACCTTGAAGCTA 300  
Db 241 ATATTGACAAAGTTGATTAATAATGAGATTAAATGATATATGATCTTACCTTGAAGCTA 300  
QY 301 AGGGTATGAGAAAACATATGATTAATGATGATTTGAGCAAGCATGAAATGAAAGCATGT 360  
Db 301 AGGGTATGAGAAAACATATGATTAATGATGATTTGAGCAAGCATGAAATGAAAGCATGT 360  
QY 361 CATCTAGAACACCTGTTCAATTTGTTGTTCTTAAACAAAGATTTATCTTCAAGCAAA 420  
Db 361 CATCTAGAACACCTGTTCAATTTGTTGTTCTTAAACAAAGATTTATCTTCAAGCAAA 420  
QY 421 TCACCTTTCAAGTTCATGACAGATCTTCTATTTCAATTAAGATTTTGGATCTTGAAG 480  
Db 421 TCACCTTTCAAGTTCATGACAGATCTTCTATTTCAATTAAGATTTTGGATCTTGAAG 480  
QY 481 CATCTAGTAAATTTTCAAGCTCAAAAGATGAAAGGCTTGGATCTTGAAGCTTC 540  
Db 481 CATCTAGTAAATTTTCAAGCTCAAAAGATGAAAGGCTTGGATCTTGAAGCTTC 540  
QY 541 AAAATTTAGTTGAGAGAGAGAGAACTATCATATGCAATGCAAGTATGAGGCCAA 600  
Db 541 AAAATTTAGTTGAGAGAGAGAGAACTATCATATGCAATGCAAGTATGAGGCCAA 600  
QY 601 GTTCTTGCAAAATTAATATATCTGCGCATGAGGAGTGCACCAAGGCTTAACTTCT 660  
Db 601 GTTCTTGCAAAATTAATATATCTGCGCATGAGGAGTGCACCAAGGCTTAACTTCT 660  
QY 661 GGAATTTGAAAAATTTGAAAGGAAATTAATCTAAAGATTAATAAGCAACAATTCATA 720  
Db 661 GGAATTTGAAAAATTTGAAAGGAAATTAATCTAAAGATTAATAAGCAACAATTCATA 720  
QY 721 TCAATTTGAGTCATGACATATATGTTAGCTTCAATTTGATGATCAATGCTTGAGCA 780  
Db 721 TCAATTTGAGTCATGACATATATGTTAGCTTCAATTTGATGATCAATGCTTGAGCA 780  
QY 781 AGAGGCCAAATTCATGATGAGTCCATGATCAAAATGCTATATTTCAATATCTGATA 840  
Db 781 AGAGGCCAAATTCATGATGAGTCCATGATCAAAATGCTATATTTCAATATCTGATA 840  
QY 841 CTATATTTGAAACAGGTATGATATTTATTTCAATTTGTTTGGATCTCAAAATGTCAGG 900  
Db 841 CTATATTTGAAACAGGTATGATATTTATTTCAATTTGTTTGGATCTCAAAATGTCAGG 900  
QY 901 CCAGAAATATTAATCTTGCTCAGGTCAATGATATAGTATGGAAGCTTGAAGATCTGAA 960  
Db 901 CCAGAAATATTAATCTTGCTCAGGTCAATGATATAGTATGGAAGCTTGAAGATCTGAA 960

```

Db      901  CCACAAATATTAATCTGGTCCAGGTCATGCTAATGATATGGAAGCTTAGATCTGGAA  960
QY      961  ATTACAGAGCTTATGTCTTAATGTTACTGTAAATGAGCCAAATTAATGCGCCGAA  1020
Db      961  ATTACAGAGCTTATGTCTTAATGTTACTGTAAATGAGCCAAATTAATGCGCCGAA  1020
QY      1021  ATGAGATTAGATCAAGACTTGGCAGGAGGAGATCTGGACAAGCTAGACATCAAAATTC  1080
Db      1021  ATGAGATTAGATCAAGACTTGGCAGGAGGAGATCTGGACAAGCTAGACATCAAAATTC  1080
QY      1081  TGAATGTGGAATGCAAGAGCTTAAGTATCCATTAATTAATGACCAAAATCAATTTGATC  1140
Db      1081  TGAATGTGGAATGCAAGAGCTTAAGTATCCATTAATTAATGACCAAAATCAATTTGATC  1140
QY      1141  GAGTTGAACCATGTATACAAACAGTTTTCAGCAGCTTCAAGTGAAGAAATGTGTATGAG  1200
Db      1141  GAGTTGAACCATGTATACAAACAGTTTTCAGCAGCTTCAAGTGAAGAAATGTGTATGAG  1200
QY      1201  ATATCAAGGGCACAAGTGCACAAGAGTGCCATAAATTTGATTGACACAAACCTTTC  1260
Db      1201  ATATCAAGGGCACAAGTGCACAAGAGTGCCATAAATTTGATTGACACAAACCTTTC  1260
QY      1261  CATTGTAAGGAATTAATGAGATATTAATTTAGAGGGGAAAGTGAAGAACCATCAG  1320
Db      1261  CATTGTAAGGAATTAATGAGATATTAATTTAGAGGGGAAAGTGAAGAACCATCAG  1320
QY      1321  AGGCTACGTGCAAAAATGTCATTTTAAACATGCTGACATGTTAGACACACACCTGACT  1380
Db      1321  AGGCTACGTGCAAAAATGTCATTTTAAACATGCTGACATGTTAGACACACACCTGACT  1380
QY      1381  CACTAGAAATTTCAAGAGATGAAGCTCTTTTGTATATTAATTTATTAATTAATTAAT  1440
Db      1381  CACTAGAAATTTCAAGAGATGAAGCTCTTTTGTATATTAATTTATTAATTAATTAAT  1440
QY      1441  TCAATATATGACGATATGATATATACAAATTAACAAATTAATTAATTAATTAATTA  1500
Db      1441  TCAATATATGACGATATGATATATACAAATTAACAAATTAATTAATTAATTAATTA  1500
QY      1501  TTAATTAATATGATGACGATGATGATTAATTAATTAATTAATTAATTAATTAATTA  1560
Db      1501  TTAATTAATATGATGACGATGATGATTAATTAATTAATTAATTAATTAATTAATTA  1560
QY      1561  GTCAAAAGTTTGAAGATGTTGATCTTTTATGTCACAAAATTAATTAATTAATTAAT  1620
Db      1561  GTCAAAAGTTTGAAGATGTTGATCTTTTATGTCACAAAATTAATTAATTAATTAAT  1620
QY      1621  G 1621
Db      1621  G 1621

```

```

RESULT 6
LEPER 1617 bp mRNA linear PLN 12-SEP-1993
LOCUS Tomato mRNA for polygalacturonase (Pg, EC 3.2.1.15).
ACCESSION X05656
VERSION X05656.1 GI:19297
KEYWORDS polygalacturonase.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 1617)
REFERENCE 1
AUTHORS Sheehy,R.E., Pearson,J., Brady,C.J. and Hatt,W.R.
TITLE Molecular characterization of tomato fruit polygalacturonase
JOURNAL Mol. Gen. Genet. 208, 30-36 (1987)
2
AUTHORS Rogers,H.J., Allen,R.L., Hamblin,W.D. and Lonsdale,D.M.
TITLE Pollen specific cDNA clones from Zea mays
JOURNAL Biochim. Biophys. Acta 1089 (3), 411-413 (1991)
MEDLINE 91316147
PUBMED 1859845

```

```

REFERENCE 3 (bases 1438 to 1438)
AUTHORS Hatt,W.R.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1997)
FEATURES
source
1. 1617
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/db_xref="taxon:4081"
/clone="F1"
47. 1420
CDS
1/note="unnamed protein product; polygalacturonase (AA
1-457)"
/codon_start=1
/protein_id="CAA29148.1"
/db_xref="GI:19298"
/db_xref="GGA:PO5117"
/db_xref="SWISS-PROT:PO5117"
/translation="VIVTQRNLSILLITITFASISTSRVNDLPEKQVYDNLBOEF
ADDFQAYLSLYLSKNIESNNIDKVDNGIKVINLSPGAKGDKYDNLAFQAMNBA
CSSRTPVQFVPEKNKNYLIKQITFSGPGRSSISVKIFGSLFASSKISDYDRLLWLF
DSYONLVGGGGTINNGOVWPPSSCKINKSLPCRPAPYLIITFMNCKNLKNNLKSIN
AQQIHTKPEKCTNNVNASNMINASAKSPNTDGVHVNTOYLOISDITIGDDICISV
SSQONVQNNITGPGHGISIGSSEAYSVTVNFAKTIIGANGRPITWOG
SGQASNIKFLNBMQVVKPIIIDNYICRVEPCIQFSAVQKVVYEVNIKTSATK
VAIKPDCSTNPFCEGIIMENINLVGSGKSEATCKNVHFNNAEHVTPHCTSLISED
EALLVNY"
1438
/note="c was g in [1]"
/citation=[1]
1465. 1470
/note="pot.polyA signal"
misc_feature 1597. 1603
/note="pot.polyA signal"
ORIGIN
Query Match 98.7%; Score 1614; DB 8; Length 1617;
Best Local Similarity 100.0%; Pred. No. 86-243;
Matches 1614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 TTCAATGACAACTTTAAACAGATACATATATATATATATATATATATATATATAT 67
Db 4 TTCAATGACAACTTTAAACAGATACATATATATATATATATATATATATATATAT 63
QY 68 TAGATTCCTCTTCATTAATTAATTTGGTTCATCAATTTCACTGTGAGAACATGT 127
Db 64 TAGATTCCTCTTCATTAATTAATTTGGTTCATCAATTTCACTGTGAGAACATGT 123
QY 128 TATTGATGACATTTATCAACAGATTATGATTAATTTCTTGAAACAGATTGGCTCA 187
Db 124 TATTGATGACATTTATCAACAGATTATGATTAATTTCTTGAAACAGATTGGCTCA 183
QY 188 TGAATTTCAAGCTATCTTCTATTTGAGCAAAAATATGAAAGCAATTAATATGTA 247
Db 184 TGAATTTCAAGCTATCTTCTATTTGAGCAAAAATATGAAAGCAATTAATATGTA 243
QY 248 CAAGTTGATTAATAATGGGATTAAGTGAATTAATGTAAGCTTTGAGAGCTAAGGTGA 307
Db 244 CAAGTTGATTAATAATGGGATTAAGTGAATTAATGTAAGCTTTGAGAGCTAAGGTGA 303
QY 308 TGGAAAAACATATGATATATGATTTGAGCAAGCATGGAATGAAGCATGTCATGAG 367
Db 304 TGGAAAAACATATGATATATGATTTGAGCAAGCATGGAATGAAGCATGTCATGAG 363
QY 368 AACACCTGTCATTTGTGCTCTTAAACAGAAATTAATCTCTCAAGCAAAATCACTT 427
Db 364 AACACCTGTCATTTGTGCTCTTAAACAGAAATTAATCTCTCAAGCAAAATCACTT 423
QY 428 TTCAGGTCAATGAGATCTCTTATTTCAATGAAGATTTTGGATCTTTAGAACATCTAG 487
Db 424 TTCAGGTCAATGAGATCTCTTATTTCAATGAAGATTTTGGATCTTTAGAACATCTAG 483
QY 488 TAAATTTCACTCAACAAATATGAAGGCTTTGGATTCCTTTTGATATGTTCAAAATTT 547

```

Db 484 TAAATTTTCAGACTACAAAGATAGAAAGCTTTGGATGCTTTTGTAGTGTCAAAATTT 543  
QY 548 AGTTGTTGAGAGAGAGAACTATCAATGCGAATGGAAGAATGTTGGCCAAAGTTCTTG 607  
Db 544 AGTTGTTGAGAGAGAGAACTATCAATGCGAATGGAAGAATGTTGGCCAAAGTTCTTG 603  
QY 608 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTG 667  
Db 604 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTG 663  
QY 668 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTT 727  
Db 664 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTT 723  
QY 728 TGAGTCATGCACTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 787  
Db 724 TGAGTCATGCACTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 783  
QY 788 AATATCTGATGAGAGTCCATGATCAATATCAATATCAATATCAATATCAATATCAATAT 847  
Db 784 AATATCTGATGAGAGTCCATGATCAATATCAATATCAATATCAATATCAATATCAATAT 843  
QY 848 TGAACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907  
Db 844 TGAACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 903  
QY 908 TATTAATGTTGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 967  
Db 904 TATTAATGTTGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 963  
QY 968 AGCTTATGTTCTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 1027  
Db 964 AGCTTATGTTCTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 1023  
QY 1028 TAGATATCAAGCTTGCGACGAGAGATCTGACAAAGCTAGCAACATCAATTTCTGAATGT 1087  
Db 1024 TAGATATCAAGCTTGCGACGAGAGATCTGACAAAGCTAGCAACATCAATTTCTGAATGT 1083  
QY 1088 GGAATGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1147  
Db 1084 GGAATGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143  
QY 1148 ACCATGTTATCAAGCTTTGAGCAGCTTCAAGTGAATATGTTGTTATGAGAAATATCAA 1207  
Db 1144 ACCATGTTATCAAGCTTTGAGCAGCTTCAAGTGAATATGTTGTTATGAGAAATATCAA 1203  
QY 1208 GGGCAAAAGTGCACAAAGGTGGCCATTAATTTGATGACGACAAACCTTCCATGTGA 1267  
Db 1204 GGGCAAAAGTGCACAAAGGTGGCCATTAATTTGATGACGACAAACCTTCCATGTGA 1263  
QY 1268 AGGAATTTATATGAGAAATTAATTTATGTTAGGGGAAAGTGGGAAACCATGAGGGCTAC 1327  
Db 1264 AGGAATTTATATGAGAAATTAATTTATGTTAGGGGAAAGTGGGAAACCATGAGGGCTAC 1323  
QY 1328 GTGCAAAATGTCATTTTAAACAATGCTGAACATGTTACACACAGCTGACCTTCACTAGA 1387  
Db 1324 GTGCAAAATGTCATTTTAAACAATGCTGAACATGTTACACACAGCTGACCTTCACTAGA 1383  
QY 1388 AATTTCAAGAGATGAAGCTCTTTTGTATTAATTTAATTTAATTTAATTTAATTTAAT 1447  
Db 1384 AATTTCAAGAGATGAAGCTCTTTTGTATTAATTTAATTTAATTTAATTTAATTTAAT 1443  
QY 1448 AATGAGATATGATATATCAATTAACAATCTATCTATCTATCTATCTATCTATCTAT 1507  
Db 1444 AATGAGATATGATATATCAATTAACAATCTATCTATCTATCTATCTATCTATCTAT 1503  
QY 1508 TAAATATGACGATGATGAGTTTAAATAGACTATCTATCTATCTATCTATCTATCTAT 1567  
Db 1504 TAAATATGACGATGATGAGTTTAAATAGACTATCTATCTATCTATCTATCTATCTAT 1563  
QY 1568 GTTTGACGATTTGATCTTTTAAATGATCAAAATTAATTAATGTTATTTATATG 1621

Db 1564 GTTTGACGATTTGATCTTTTAAATGATCAAAATTAATTAATTAATTTATATG 1617  
RESULT 7  
LOCUS 101809 1624 bp ss-DNA linear PAT 21-MAY-1993  
DEFINITION Sequence 1 from Patent US 4801540.  
ACCESSION 101809  
VERSION 101809.1 GI:269802  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1624)  
AUTHORS Hiatt, W.R., Sheehy, R.E., Shewmaker, C.K., Kridl, J.C. and Knauf, V.  
TITLE Pg. gene and its use in plants  
JOURNAL Patent: US 4801540-A 1 31-JAN-1999;  
CALGENE, Inc.; Davis, CA  
FEATURES  
location/Qualifiers  
source 1..1624  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Query Match 98.6%; Score 1613.2; DB 6; Length 1624;  
Best Local Similarity 99.8%; Pred. No. 1.1e-242;  
Matches 1615; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 8 TTCAATAGACAAAGTTTAAACCATATACATATATATATATATATATATATATATATAT 67  
Db 4 TTCAATAGACAAAGTTTAAACCATATACATATATATATATATATATATATATATATAT 63  
QY 68 TAGATATGCTGCTGATATATATTTTGTGCTGATCAATTTCAATTTGAGAGAGATGT 127  
Db 64 TAGATATGCTGCTGATATATATTTTGTGCTGATCAATTTCAATTTGAGAGAGATGT 123  
QY 128 TATGATGACAAATTTATTCAAACAAAGTTATGATATATATTTCTTGAAACAAATTTGCTCA 187  
Db 124 TATGATGACAAATTTATTCAAACAAAGTTATGATATATATTTCTTGAAACAAATTTGCTCA 183  
QY 188 TGATTTTCAAGCTTATCTTTCTTATTTAGAGCAAAATTTGAAAGCAACATTAATTTGA 247  
Db 184 TGATTTTCAAGCTTATCTTTCTTATTTAGAGCAAAATTTGAAAGCAACATTAATTTGA 243  
QY 248 CAAGTTGATTAATGAGGATTAAGATGATTAATGATCTTGAAGCTTGAAGCTTGAAGGATGA 307  
Db 244 CAAGTTGATTAATGAGGATTAAGATGATTAATGATCTTGAAGCTTGAAGCTTGAAGGATGA 303  
QY 308 TGAACAAACATATATATATTTGATTTGACCAAGCATGAAATGAAGCATGTTCACTAG 367  
Db 304 TGAACAAACATATATATATTTGATTTGACCAAGCATGAAATGAAGCATGTTCACTAG 363  
QY 368 AACACCTGTTCAATTTGTTGTTCTTAACAAACAAAGATTTATCTTCAAGCAATATCACTT 427  
Db 364 AACACCTGTTCAATTTGTTGTTCTTAACAAACAAAGATTTATCTTCAAGCAATATCACTT 423  
QY 428 TTCAGTTCATGACAGATCTTCTATTTCAATTAAGATTTTGGATCCTTGAAGCATAG 487  
Db 424 TTCAGTTCATGACAGATCTTCTATTTCAATTAAGATTTTGGATCCTTGAAGCATAG 483  
QY 488 TAAATATTCAGACTTCAAAAGATGAAGGCTTGGATGCTTTGATAGTGTTCAAATTT 547  
Db 484 TAAATATTCAGACTTCAAAAGATGAAGGCTTGGATGCTTTGATAGTGTTCAAATTT 543  
QY 548 AGTTGTTGAGAGAGAGAACTATCAATGCGAATGGAAGAATGTTGGCCAAAGTTCTTG 607  
Db 544 AGTTGTTGAGAGAGAGAACTATCAATGCGAATGGAAGAATGTTGGCCAAAGTTCTTG 603  
QY 608 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTG 667  
Db 604 CAAATATATATATCACTGCGATGAGAGGATGACCAACGCGCTTAACTTCTTGGAATTG 663  
QY 668 CAAATATTTGAAAGTGAATTAATCTAAGAGTAAATATGACAAACATTCATATCAATTT 727

```

Db      664 CAAAAATTGAAAGTGAATTAATCTAAGAGTAAAGCAACAAATTATATCAAAATT 723
      728 TGAAGTATGCACTAATTTGTAGCTTCAAAATTTGATGATCAATAGCTTCAGCAAGAGCCC 787
      724 TGAGTCATGCACTAATTTGTAGCTTCAAAATTTGATGATCAATAGCTTCAGCAAGAGCCC 783
      788 AAATAGTATGAGTGCATGATCAATACTCAATATATTTCAAAATATCTGATCTATAT 847
      784 AAATAGTATGAGTGCATGATCAATACTCAATATATTTCAAAATATCTGATCTATAT 843
      848 TGAACACAGTGAATTTGTATTCATTTGTTCTGTGATCTCAAAATGTGAGGCCACAA 907
      844 TGAACACAGTGAATTTGTATTCATTTGTTCTGTGATCTCAAAATGTGAGGCCACAA 903
      908 TATTACTGTGTGTCAGAGTCAATGATTAAGTATTTGGAAGCTTGAAGCTTGAAGTTCAGA 967
      904 TATTACTGTGTGTCAGAGTCAATGATTAAGTATTTGGAAGCTTGAAGCTTGAAGTTCAGA 963
      968 AGCTTATGTGTCTAATGTTAATCTGTAAATGAAGCCAAATTTATCGTGCCGAAAATGAGT 1027
      964 AGCTTATGTGTCTAATGTTAATCTGTAAATGAAGCCAAATTTATCGTGCCGAAAATGAGT 1023
      1028 TAGGATCAAGACTTTGGCAGGAGAGATCTGCAACAGCTAGCAACATCAAAATTTCTGAATG 1087
      1024 TAGGATCAAGACTTTGGCAGGAGAGATCTGCAACAGCTAGCAACATCAAAATTTCTGAATG 1083
      1088 GGAATATGCAAGCGTTAAGTATCCCAATTAATATAGACCAAACTATTTGTGATCGAGTTGA 1147
      1084 GGAATATGCAAGCGTTAAGTATCCCAATTAATATAGACCAAACTATTTGTGATCGAGTTGA 1143
      1148 ACCATGATATCAACAGTTTTCAGCAGTTCAAGTGAAGAAAATGTGTGTATGAGATATCAA 1207
      1144 ACCATGATATCAACAGTTTTCAGCAGTTCAAGTGAAGAAAATGTGTGTATGAGATATCAA 1203
      1208 GGGCACAAGTGCACAAAGGTGGCCATTAATTTGATGACACCAACTTTCCATGTGA 1267
      1204 GGGCACAAGTGCACAAAGGTGGCCATTAATTTGATGACACCAACTTTCCATGTGA 1263
      1268 AGGAATATATATGAGATTAATTAATTTAGTAGGGGAAAGTGAAGAACCATCAGAGGTAC 1327
      1264 AGGAATATATATGAGATTAATTAATTTAGTAGGGGAAAGTGAAGAACCATCAGAGGTAC 1323
      1328 GTGCACAAATGTCATTTTAACATGCTGAACATGTTACACACACTGCATCTCATTAGA 1387
      1324 GTGCACAAATGTCATTTTAACATGCTGAACATGTTACACACACTGCATCTCATTAGA 1383
      1388 AATTTCAAGAGATGAGAGCTCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 1447
      1384 AATTTCAAGAGATGAGAGCTCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 1443
      1448 ATAGCAGATATGATATATCAATTAACAAATCTATCTATGATGATGATTAATTAAT 1507
      1444 ATAGCAGATATGATATATCAATTAACAAATCTATCTATGATGATGATTAATTAAT 1503
      1508 TAAATATGATGAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1567
      1504 TAAATATGATGAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1563
      1568 GTTTGACGATTTGATTTTAAATGATCAAAATTAATTAATTAATTAATTAATTAAT 1625
      1564 GTTTGACGATTTGATTTTAAATGATCAAAATTAATTAATTAATTAATTAATTAAT 1621

```

RESULT 8  
 CPAS05750  
 LOCUS CPAS05750 599 bp mRNA linear PLN 04-SEP-2003  
 DEFINITION Carica papaya partial mRNA for putative polygalacturonase (pg  
 gene).  
 ACCESSION AF505750  
 VERSION AF505750.1 GI:34481836  
 KEYWORDS pg gene, polygalacturonase.  
 SOURCE Carica papaya (papaya)

```

ORGANISM Carica papaya
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLES rosids; eurosids II; Brassicales; Caricaceae; Carica.
JOURNAL Saiprasad, G.V.S. and Lalitha, A.
REFERENCE Isolation, cloning and characterization of polygalacturonase gene
AUTHORS from ripening papaya fruit (Carica papaya cv. Singapore pp-1)
TITLES Unpublished
JOURNAL (bases 1 to 599)
AUTHORS Saiprasad, G.V.S.
JOURNAL Direct Submission
TITLE Submitted (13-AUG-2002) Saiprasad G.V.S., Biotechnology, Indian
JOURNAL Institute of Horticultural Research, IIHR, Hesaraghatta lake post,
JOURNAL Bangalore, Karnataka, INDIA
FEATURES
source
1..599
/organism="Carica papaya"
/mol_type="mRNA"
/cultivar="Singapore pp-1"
/db_xref="taxon:3649"
/life_type="fruit"
1..599
/gene="pg"
<1..>599
/gene="pg"
/codon_start=2
/product="putative polygalacturonase"
/protein_id="CAD44275.1"
/db_xref="GI:34481837"
/translation="EQAMNDCSRPVPQFVPERKNNYLKQITPSPGSSISVKLF
GSLEASKISDYDRRLMIFDSVQNLVVGGGTIGNGVWVPSGCKIKSLPCRD
PALTFPWNCKNLKVNLMKSAQOIHKEPSCNNVVASNIMINIASAKSPMTDGVHSN
TQYIQISDRIIGDQICISVSGSNVQATNIRICPGHG"
ORIGIN
Query Match 36.6%; Score 599; DB 8; Length 599;
Best Local Similarity 100.0%; Pred. No. 3,9e-84;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db      335 TGAGCAGCATGGAATGAGAACATGTTCACTATTAACACCGTGCATTTGGTCTCTAA 394
      1 TGAGCAGCATGGAATGAGAACATGTTCACTATTAACACCGTGCATTTGGTCTCTAA 60
      395 AAACAGAAATTAATCTCTCAAGCAATCACTTTTCAGTGCATGAGATCTTCTATTTTC 454
      61 AAACAGAAATTAATCTCTCTCAAGCAATCACTTTTCAGTGCATGAGATCTTCTATTTTC 120
      455 AGTAAAGATTTTGGATCCTTAAGAGATCTAGTAAATTTTCAGACTACAAAGATTAAG 514
      121 AGTAAAGATTTTGGATCCTTAAGAGATCTAGTAAATTTTCAGACTACAAAGATTAAG 180
      515 GCTTTGGATGCTTTTGGATGCTTGAATTTAGTTGTTGAGAGAGAGAACTATCAA 574
      181 GCTTTGGATGCTTTTGGATGCTTGAATTTAGTTGTTGAGAGAGAGAACTATCAA 240
      575 TGCCAAATGACAGATATGATGAGCAAGTTCTTGCAAAATTAATTAATCACTGCCATGAG 634
      241 TGCCAAATGACAGATATGATGAGCAAGTTCTTGCAAAATTAATTAATCACTGCCATGAG 300
      635 GGATGACCAACAGCGCTTAACTTTCTGGAATTTGCAAAATTTGAAGTGAATATCTTAA 694
      301 GGATGACCAACAGCGCTTAACTTTCTGGAATTTGCAAAATTTGAAGTGAATATCTTAA 360
      695 GAGTAAATATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 754
      361 GAGTAAATATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
      755 AAATTTGATGATCAATGCTTCAGCAAAAGAGCCCAATTAATTAATTAATTAATTAAT 814
      421 AAATTTGATGATCAATGCTTCAGCAAAAGAGCCCAATTAATTAATTAATTAATTAAT 480
      815 TACTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 874

```

Db 481 TACTCAATATATGCAATATCTGATCTATTTATGGAACAGGTGATGATGATTTCAAT 540  
Qy 875 TGTTCCTGATCTCAAAATGTCAGGCCAACAATATTAATTGTCAGTCATGTA 933  
Db 541 TGTTCCTGATCTCAAAATGTCAGGCCAACAATATTAATTGTCAGTCATGTA 599

RESULT 9  
LESS05947 602 bp mRNA linear PIN 04-SEP-2003  
LOCUS Lycopersicon esculentum partial mRNA for putative polygalacturonase  
DEFINITION (pg gene) .  
ACCESSION AJ505947  
VERSION AJ505947.1 GI:34481842  
KEYWORDS pg gene; polygalacturonase.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
REFERENCE 1 Saiprasad,G.V.S.  
AUTHORS Isolation, cloning and characterization of polygalacturonase gene  
TITLE from fruit tissue of Lycopersicon esculentum cv. Arka vikas  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 602)  
AUTHORS Saiprasad,G.V.S.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-2002) Saiprasad G.V.S., Biotechnology, Indian  
Institute of Horticultural Research, IHR, Hesaraghatta Lake post,  
Bangalore, Karnataka, INDIA  
FEATURES  
source Location/Qualifiers  
1..602  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="Arka vikas"  
/db\_xref="taxon:4081"  
/tissue\_type="fruit"  
1..602  
/gene="pg"  
/gene="pg"  
/gene="pg"  
/codon\_start=2  
/evidence="experimental"  
/product="putative polygalacturonase"  
/protein\_id="CAD44521.1"  
/db\_xref="GI:34481843"  
/translation="EQANNEACSSRTPVQFVVPKKNKYLKQITFSGFCRSSISEKIF  
GSLSSKISDYKDRLLDCLKSYQNLVVGSGTINNGOVMPSSCKIKRSLPCRD  
APTALTTFWNCNKLKVNKSKNMQOIHKEFGCTNVVSNLMINASAKSPMTDGVHS  
NTQYIQISDTITIGTGDCISIVSSGQNVQAINITGPHG"

ORIGIN  
Query Match 34.8%; Score 570; DB 8; Length 602;  
Best Local Similarity 97.8%; Pred. No. 1.3e-79;  
Matches 589; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

Qy 335 TGAGCAGCAGTGAATGAAGATGTTCTATGAAACACCTGTTCAATTTGGTCTCTAA 394  
Db 1 TGAGCAGCAGTGAATGAAGATGTTCTATGAAACACCTGTTCAATTTGGTCTCTAA 60  
Qy 395 AAACAGAGTATCTCTCAAGCAAAATCACCTTTCAAGTCATGCAATCTTCTATTC 454  
Db 61 AAACAGAGTATCTCTCAAGCAAAATCACCTTTCAAGTCATGCAATCTTCTATTC 120  
Qy 455 AGTAAGATTTTGGATCCTTAGAAGCATCTAGTAAATTTCAAGCTACAAAGATAGAAG 514  
Db 121 AGAAAGATTTTGGATCCTTAGAAGCATCTAGTAAATTTCAAGCTACAAAGATAGAAG 180  
Qy 515 GCTTTGGATT--GCTTTGATAGTGTTCAAAATTTAGTGTGGAGAGAGAACTAT 571  
Db 181 GCTTTGGATGCTTTGAAAAGTGTTCAAAATTTAGTGTGGAGAGAGAACTAT 240

Qy 572 CAATGCAATGACAGATATGTTGGCCAAAGTTCTTGCAAAAATTAATTAATCACTGCCAG 631  
Db 241 CAATGCAATGACAGATATGTTGGCCAAAGTTCTTGCAAAAATTAATTAATCACTGCCAG 300  
Qy 632 CAGGATGACCAACAGGCTTAACTCTGGAATTCGAAAAATTTGAAAGTAATATCT 691  
Db 301 CAGGATGACCAACAGGCTTAACTCTGGAATTCGAAAAATTTGAAAGTAATATCT 360  
Qy 692 AAAGATTAATAATGACACAAATTCATATCAAAATTTGAGTCATGCACTAATGTTGAC 751  
Db 361 AAAGATTAATAATGACACAAATTCATATCAAAATTTGAGTCATGCACTAATGTTGAC 420  
Qy 752 TTCAATTTGATGATCAATGCTTCAGCAAAAGGCCCAATCACTAGTGGATGCAATGATC 811  
Db 421 TTCAATTTGATGATCAATGCTTCAGCAAAAGGCCCAATCACTAGTGGATGCAATGATC 480  
Qy 812 AATATCTCAATATATTTCAAAATATCTGATATTAATTTGAAACAGGTGATGATTTTC 871  
Db 481 AATATCTCAATATATTTCAAAATATCTGATATTAATTTGAAACAGGTGATGATTTTC 540  
Qy 872 AATGTTTCTGATCTCAAAATGTCAGGCCAACAATATTAATTGTCAGTCATG 931  
Db 541 AATGTTTCTGATCTCAAAATGTCAGGCCAACAATATTAATTGTCAGTCATG 600  
Qy 932 TA 933  
Db 601 TA 602

RESULT 10  
AF152758 1541 bp mRNA linear PIN 19-MAY-2000  
LOCUS Actinidia chinensis clone CKPgA-3 polygalacturonase A (pGa) mRNA.  
DEFINITION complete cds.  
ACCESSION AF152758  
VERSION AF152758.1 GI:7959982  
KEYWORDS  
SOURCE  
ORGANISM Actinidia chinensis  
Actinidia chinensis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; Ericales; Actinidiaceae; Actinidia.  
REFERENCE 1 (bases 1 to 1541)  
AUTHORS Wang,Z.Y., Macrae,E.A., Wright,M.A., Bolitho,K.M., Ross,G.S. and  
Ackinson,R.G.  
TITLE Polygalacturonase gene expression in kiwifruit: relationship to  
fruit softening and ethylene production  
JOURNAL Plant Mol. Biol. 42 (2), 317-328 (2000)  
MEDLINE 20252524  
PUBMED 10794531  
REFERENCE 2 (bases 1 to 1541)  
AUTHORS Ackinson,R.G.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAY-1999) Plant Development, HortResearch, Private  
Bag 92169, Auckland, New Zealand  
FEATURES  
source Location/Qualifiers  
1..1541  
/organism="Actinidia chinensis"  
/mol\_type="mRNA"  
/db\_xref="taxon:3625"  
/clone="CKPGA-3"  
/tissue\_type="fruit pericarp"  
1..1541  
/gene="pGa"  
1..1392  
/gene="pGa"  
/gene="pGa"  
/codon\_start=1  
/product="polygalacturonase A"  
/protein\_id="AAF71160.1"  
/db\_xref="GI:7959983"  
/translation="MALQRHFFQVITLLIPSFLLGYTSAVHEDPPHDYHLEBYGD  
FKAVPSYITITGNDPFGSMHENGIFGLIRVDYDMRVLVLSKTVVNDVDFGAKDGS

DDTKAFKAKVAVCSSTSSAVLLVPQKNLYVREITSGECKSLTMOIYGTLEASDR  
 SYSKSKGRHMLVDSVONLRVEGGGIIINGOLWONSCKTNTKLPCKAPALTLTK  
 SKHIVKLNKLENAQOIHSFDCNVAVSNLMTTPSPNTDGLHVTGQTHIS  
 CVLETGDDICISIVSGKVRVNDITCGPHGISIGSLGXNSNAHSDVYNARKICG  
 TTVGKIKTWOGSGSASNIKFPONVHNHVENPIIIDONVCDQDKCOEBSAVCYKN  
 IFYKNTKGTCSANVAITFDCKSRFPQGIYLEDVDEIEBGAAAKALCNVELSEYGV  
 VSPHCEBGEBSAS"

## ORIGIN

Query Match 33.6%; Score 549.2; DB 8; Length 1541;  
 Best Local Similarity 70.0%; Pred. No. 1.9e-76;  
 Matches 785; Conservative 0; Mismatches 328; Indels 9; Gaps 3;

270 AAAAGTATTAATGTTACTTACCTTTGAGCTTAAGGCTGATGGAACAAATGATTAATTT 329  
 262 AAAAGTATTAATGTTACTTACCTTTGAGCTTAAGGCTGATGGAACAAATGATTAATTT 318  
 330 GCATTGAGCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389  
 319 GCTTTGAGCAAGGCTTGAAGGCAAGGCTTGAAGGCAAGGCTTGAAGGCAAGGCTTGAAG 378  
 390 CCTAAAGCAAGATTAATCTTCTCAAGCAATCACTTTGAGCTGATGATGATGATGATGAT 449  
 379 CCCCAG--AAGAACTATCTTGTAGGCCAATAACATTCAGGCCCATGCAATCTGAT 435  
 450 ATTTCGCTAAAGATTTTGGATCCTTGAAGCATCTAGTAAATTTTCAAGCTA--CAAA 506  
 436 CTCACATGACAGATTTATGAGCACTGGAAGCATCCGATGATGATGATGATGATGATGAT 495  
 507 GATAGAAGGCTTGGATGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 566  
 496 GATGAGAGACACTGGCTTGTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 555  
 567 ACTATCAATGSCATGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 626  
 556 ACCATCAATGSCATGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615  
 627 CCATGCAAGGATGACCAACGAGCTTAACTTCTGGAATTTGAAAAATTTGAAAGTGAAT 686  
 616 CCTCGAAGGACGCTCCCAAGGCTTCAATTTTAAAGAGCAAGCATGATGATGATGATGATGAT 675  
 687 AACTTAAAGATGAAAAATGACAAACAAATTCATATCAATTTGATGATGATGATGATGATGAT 746  
 676 AACTTAAAGATGAAAAATGACAAACAAATTCATATCAATTTGATGATGATGATGATGATGAT 735  
 747 GATGCTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 806  
 736 CAAGCATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 795  
 807 GATCAAAATGATCAATTAATTTCAATATCTGATATCTATTAATTTGAAACAGGTGATGATGAT 866  
 796 GTCACCGGATCCCAAAACATTCATTTTCAAGCTGCGTCAAGACAGGTGATGATGATGATGAT 855  
 867 ATTTCATTTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 926  
 856 ATTTCATTTGTAAGCGGGGCTCAAGAGGCTGCGGCTCAATATATTAACCTGTGACCAAGGT 915  
 927 CATGATATAGTATGGAAGCTTAGATCTGGAATTTGGAATTTGGAATTTGATGATGATGATGAT 986  
 916 CATGATATAGTATGGAAGCTTAGATCTGGAATTTGGAATTTGGAATTTGATGATGATGATGAT 975  
 987 ACTGTAATGAAAGCCAAAATTATCGGTGCCGAAATGGAATTTGATGATGATGATGATGATGAT 1046  
 976 GTGGTGAATGAGCTTAAGCTTTGTGGAACATCAATTTGATGATGATGATGATGATGATGATGAT 1035  
 1047 GGAAGATCTGGAAGCTTAGATCTGGAATTTGGAATTTGGAATTTGATGATGATGATGATGATGAT 1106  
 1036 GGGGATCTGGAAGCTTAGATCTGGAATTTGGAATTTGGAATTTGATGATGATGATGATGATGAT 1095  
 1107 TATCCCATATTAATGACCAAACTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1166  
 1096 AACCCATCATTTATACCAAAATTTACTGTGACCAAGACAAACCTTGTCTAAGACAGAGC 1155

QY 1167 TCAGCAGCTTCAAGTGAATAATGTTGTTATGAGATATCAAGGCAACAGCTGCACAAAG 1226  
 DB 1156 TCAGCAGCTTCAAGTGAATAATGTTGTTATGAGATATCAAGGCAACAGCTGCCTCCAAAT 1215  
 QY 1227 GTGGCCATTAATTTGATTTGACGACAACTTCCATGATGATGATGATGATGATGATGATGAT 1286  
 DB 1216 GTGGCCATTAATTTGATTTGACGACAAAGGTTCCATGATGATGATGATGATGATGATGAT 1275  
 QY 1287 ATAAATTTAGTGGGGAAGATGGAACCATCAGAGGCTACGTGCAAAAATGTCATTTT 1346  
 DB 1276 GTTGATCTTGAAGTATGAGAGGCTGACAGCAAGAAAGCTTTGTCACAAATGTCGAATC 1335  
 QY 1347 AACATGCTGAACATGTTACACCACTGCACTTCACTAGAA 1388  
 DB 1336 AGTGAACAGAGGTTGTTTACACCACTGCTCCAGAGAGAA 1377

## RESULT 11

AY043233

LOCUS 1595 bp mRNA linear PLN 20-MAY-2003  
 DEFINITION Vitis vinifera polygalacturonase mRNA, complete cds.  
 ACCESSION AY043233  
 VERSION AY043233.1 GI:15081599  
 KEYWORDS

## SOURCE

ORGANISM

Vitis vinifera  
 Vitis vinifera  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; Vitaceae; Vitis.

## REFERENCE

1 (bases 1 to 1595)  
 Nunan,K.J., Davies,C., Robinson,S.P. and Fincher,G.B.  
 Expression patterns of cell wall-modifying enzymes during grape  
 berry development  
 Planta 214 (2), 257-264 (2001)

## JOURNAL

MEDLINE

2158619

PUBMED

11800390

REFERENCE

2 (bases 1 to 1595)

Fincher,G.B.

JOURNAL

Direct Submision

Submitted (02-JUL-2001)

Depot. Plant Science, University of

Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia

3 (bases 1 to 1595)

Nunan,K.J., Davies,C. and Robinson,S.P.

JOURNAL

Direct Submision

Submitted (02-JUL-2001)

Plant Industry, CSIRO, Hartley Grove,

Adelaide, SA 5064, Australia

FEATURES

location/Qualifiers

1..1595

/organism="Vitis vinifera"

/mol\_type="mRNA"

/cultiivar="Shiraz"

/db\_xref="taxon:29760"

55..1389

/codon\_start=1

/product="polygalacturonase Fgl"

/protein\_id="AAK81876.1"

/translation="MSLORRFPVCLIVFLSPCYRSLDASPPDYLTGEASGYDSOAN

HHHSRTGSRRTENMPMKRFSGKTSGLASVYKMNYYGAKGDSQDAPEAKKAKA

ACSSPQSVLVPENKNYLLKPIPTQSPKSSITVQVYGVQSTDSRATSNDRTHWLI

FENVONLAVQGGGTINGNGKTWENSCKNYDLPCKGAPLALTFYCKNLAVNLKIQ

NAQQMIVSEFEKCVQASGLITTPAGNSPNTDGIHVSYTKIIQISSVIGTGDICISI

VSGSQVYVNGITCGPHGISIGSLGXNSNAHSDVYNARKICG

GSGSASNIKFPONVHNHVENPIIIDQYCDQDKPCQSSAVOVQVIVYONIRGTSSS

KEAHSIDCAKRPCCGIILRDIDIKYGGKAKAKAVCSNARVYTMGDSVPCA"

## ORIGIN

Query Match 32.1%; Score 525.2; DB 8; Length 1595;  
 Best Local Similarity 65.6%; Pred. No. 1e-72;  
 Matches 815; Conservative 0; Mismatches 418; Indels 9; Gaps 3;

QY 268 TTTAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 327



Db 278 TTAATAGTGTAAAGTATATTATATGAGCTAAAGTGATG---AAGTATGCCACAG 334  
Qy 328 TTGATTTGACAGACATGATGATGATGTTTATCTAGAACACCTGTTCAATTTGGG 387  
Db 335 AGGATTCATGAAGAGCTTGAAGGAGCTTGTATCTCCAGAGATCTGTCTTA---GTGG 391  
Qy 388 TTCCATAAAGACAGATATATCTTCCAGCAAAATCACCTTTCCAGTCCATGACAGATCTT 447  
Db 392 TGCCCAAAAACAGAGCATCTCTCTTAAGCCATTTACATTTCCAGAGTCTCTGTAAATCA 451  
Qy 448 CTATTTTCAGTAAAGATTTTGGATCTTTAGAGCATCTAGTAAATTTCAAGCTAC--A 504  
Db 452 GTATTTACAGTGCAGATATATGGAACCTGTCCAGGCACTACTGATGATGCCCTTACAGTA 511  
Qy 505 AAGATTAAGAGGCTTTGGATGCTGTTGTATGTTCAAAATTTAGTTGTTGAGAGAGAG 564  
Db 512 ACGATATACCCACCTGCTTATCTTTGAAAATGTCCAAAATTTAGCGGTTCAAGGTGTGTG 571  
Qy 565 GAACTATCAATGCAATGAGACAAAGATGATGAGCCAAAGTTCTTGGCAAAATTAATTAATCAC 624  
Db 572 GCACCATTAATGGAATGGAAGAGAGCTGTGGAAAACCTCTTGCAGAGTGAATATGATC 631  
Qy 625 TGGCATGACGAGGATGACCAACGGCTTAACTCTTGTGAATTTGCAAAAATTTGAAAGTGA 684  
Db 632 TTCCCTTGCAAGGAGTGCACCAACGGCTTAACTCTTAACTGCAAGAACTTGGCAGTGA 691  
Qy 685 ATATCTTAAGAGTAAATGACAAACAAATTCATGAATTTGAGTCATGCACTATG 744  
Db 692 AGAACCTCAAGATCCAAATGCGCAGCAAAATGCAATGATGTTCTTTGAAAATAGCGTGGGTG 751  
Qy 745 TTGAGCTTCAAAATTTGATGATCAATGCTTCAGCAAAAGGCCAAATACTGATGAGTCC 804  
Db 752 TTCAAGCTCGGGGTCTCCATTACTGACCCGGGGAACAGTCCCAACTGATGAAATCC 811  
Qy 805 ATGATCAAAATCTCAATATATTCOAATATCTGATCTATTAATTTGGAACAGGTGATGTT 864  
Db 812 ATGTAGATTAACCAAAATTTATTCAGATTTCAAGCTCGGTTATTTGGAACAGGTGATGACT 871  
Qy 865 GTATTTCAATGTTTTCGATCTCAAAATTTGCAAGGCGCAAAATTTACTGTGTGCTCAG 924  
Db 872 GTATTTCAATGTTGATGATCTCCAGAAAGTGCAGATGATGATCACTGTGACAGAG 931  
Qy 925 GTCATGTATAGTATTTGGAAGCTTTAGATCTGCAAAATTTCAAGACTTATGTCTAATG 984  
Db 932 GCCATGGAATCAGCATTTGGAAGCTTAGATCTGAGATTTGAGAGCTCATGTTGCGATG 991  
Qy 985 TTACTGTAAATGAAGCCAAATTTATCGGTGCCAAAATGAGATTGATCAAGACTTGGC 1044  
Db 992 TTACAGTGAACGGGCGCAACGCTTTCGGGAAACCAAAATGAGATTGAAATCAAGACATGGC 1051  
Qy 1045 AGGAGAGATCTGACAGCTAGCAACATCAAAATTTCTGAATTTGGAATGCAAGACGTTA 1104  
Db 1052 AGGAGAGGCTGGAAGCGCGAAGCAACATCAAAATTTCAAAATGTAATGCAATTAACGTTG 1111  
Qy 1105 AGTATCCCATATTTATAGACCAAAATCTATTTGATCGATGGAACCATGTATACAAACAGT 1164  
Db 1112 AGAACCCATATATATATAGCAAAAGTACTGTGACCAAGCAACATGCAAGTACAGTA 1171  
Qy 1165 TTTGACGAGTTCAAGTGAATAATGTGTGTATGAGATATCAAGGCGCAAGTGCACAA 1224  
Db 1172 GCTCGGCTGTTCAGATGCAAAATGTGTGTACCAAGATATTAAGGAACAAAGTTCTTCTGGA 1231  
Qy 1225 AGGAGGCGCATAAATTTGATTTGAGCAGCAAACTTCCATGGAAGAGATTTTAATGAGGA 1284  
Db 1232 AGGAGGCTATATCGCTTATTTGCAAGTGCAGAGTTTCCATGTCAGAGATTTCTGCTGCGAG 1291  
Qy 1285 ATATTAATTTTATAGGGAAGAGTGAATAACATCAAGAGCTTACGCAAAATATGTCATT 1344  
Db 1292 ACATTTGATATTAAGATTTGAGAGAGGAAGGAGCAAGCAAGCTGTATGACAGCATGCAAG 1351  
Qy 1345 TTAACAATGCTGAACATGTTTACCAACACTGCACTTCACTAGAAATTTCAAGAGATGAAG 1404  
Db 1352 TGACTGTATATGGGAGATGTGTCTCCAAATTTGCGCTGAGAAATGATGGAAGCTTAGGATT 1411

Qy 1405 CTCCTTTGATATATTTATTTATTTATGATGATCTTCATATATATAGAGATATGATATA 1464  
Db 1412 ATGTGTTGATGATCTCTGATCAACACTTACATATATATATTTTGTATATATATATAC 1471  
Qy 1465 TCACATTAACAAATCTATATCTATGTTGATTTGATTAATTAATTA 1506  
Db 1472 AGTACTGAAAATATATATATGATGTTGTGCATATTTGTTATATCA 1513

RESULT 12  
TOMPGAAA  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT

FEATURES  
source  
prim\_transcript  
CDS

1..7456  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4081"  
/note="polygalacturonase mRNA and introns"  
join(1479..1757,2416..2547,3327..3491,3696..3716,  
4260..4467,4567..4648,5602..5710,6139..6255,6788..7048)  
/note="polygalacturonase"  
/codon\_start=1  
/protein\_id="AAA34178.1"  
/db\_xref="GI:170473"  
/translation="MVIQRNSILDLIIIFASSISTCRSNVIDNLFRQVYNILIEOF  
AHDFOAYLSTYKNIENNDKDKNGIKIVNLISGAKGDGYDIAEQAWNEA  
CSSRTPQVQPVYRKXNYLKOITSGGCRSGISYKIGSLFASSKISYKRPRLIAF  
DSVDNLVVGSGGTTINGEQWPPSSCKINSLPRDAPFALTTPNCKKLKXNNLSKN  
AQOIHKEFSCNTVNASNLIMNASAKSPNDGVASNTQYI01DITIIIGTDDCISIV  
SGSONVQATNTTCGPHGIGISIGLSGNSSEYVENVNTYKIIIGAEVGRIKWOG  
SGQASNIKFLNEMQDYKYPILIDQNTCDRVEPCIOQSAVQKXNVYENIKGTSATK  
VAIKPDCSTNPRCGIIMENINLVGSEKSPSEATCKVHFNNAHVTHCHLSLESD  
EALYNY"  
<1479..1757  
/note="polygalacturonase"  
/number=1  
1758..2415  
/note="polygalacturonase intron A"  
2416..2547  
/number=2  
2548..3326  
/note="polygalacturonase intron B"  
3327..3491  
/note="polygalacturonase intron C"  
3492..3695  
/note="polygalacturonase intron C"  
3696..3716  
/note="polygalacturonase intron D"  
3717..4259  
/note="polygalacturonase intron E"  
4260..4467  
/note="polygalacturonase intron E"  
4468..4566  
/number=6  
4567..4648  
/number=6  
4649..5601  
intron



/note="polygalacturonase intron F (no splice consensus); putative"  
exon 5602..5710  
/number=7  
intron 5711..6138  
/note="polygalacturonase intron G"  
exon 6139..6255  
intron 6256..6787  
/note="polygalacturonase intron H"  
exon 6788..>7048  
/note="polygalacturonase"  
/number=9

## ORIGIN

Query Match 28.2%; Score 462; DB 8; Length 7456;  
Best Local Similarity 100.0%; Pred. No. 5,3e-63;  
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1161 CAGTTTCAGCAGTTCAAGTGAATAATGTGTATGAGAAATCAAGGGCACAAGTGCA 1220  
|||  
DB 6785 CAGTTTCAGCAGTTCAAGTGAATAATGTGTATGAGAAATCAAGGGCACAAGTGCA 6844  
|||  
QY 1221 ACAAAGTGGCCATAAATTTGATGCGACGACAACCTTCCATGTAAGGAATTATATG 1280  
|||  
DB 6845 ACAAAGTGGCCATAAATTTGATGCGACGACAACCTTCCATGTAAGGAATTATATG 6904  
|||  
QY 1281 GAGATATATAATTAGTAGGGGAAAGTGAACATCAAGAGGCTACGTGCAAAATATGTC 1340  
|||  
DB 6905 GAGATATATAATTAGTAGGGGAAAGTGAACATCAAGAGGCTACGTGCAAAATATGTC 6964  
|||  
QY 1341 CATTTTAACAATGCTGAACATGTTACACCACTGCACTTCACTAGAAATTTCAAGGAT 1400  
|||  
DB 6965 CATTTTAACAATGCTGAACATGTTACACCACTGCACTTCACTAGAAATTTCAAGGAT 7024  
|||  
QY 1401 GAAAGCTTTGTAT 1460  
|||  
DB 7025 GAAAGCTTTGTAT 7084  
|||  
QY 1461 TATATCAATATAACCAATCTATATCTATGATTTGAATAATTTATTTATATATGCGA 1520  
|||  
DB 7085 TATATCAATATAACCAATCTATATCTATGATTTGAATAATTTATTTATATATGCGA 7144  
|||  
QY 1521 TTGAAGTTTATTAAGACTATATGATTTTCTATTTTCTAGTCAAAAGTTTGCGATTGT 1580  
|||  
DB 7145 TTGAAGTTTATTAAGACTATATGATTTTCTATTTTCTAGTCAAAAGTTTGCGATTGT 7204  
|||  
QY 1581 ACTTTTAAATGTACAAATAATTAATAAATGTTATTTATATATGA 1622  
|||  
DB 7205 ACTTTTAAATGTACAAATAATTAATAAATGTTATTTATATGA 7246  
|||

RESULT 13  
LEPOLYGA 7456 bp DNA linear PLN 09-MAY-1995  
LOCUS  
DEFINITION Tomato gene for cell wall degrading polygalacturonase.  
ACCESSION X14074  
VERSION X14074.1 GI:19305  
KEYWORDS polygalacturonase.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanales; Solanales; Lycopersicon.  
1 (bases 1 to 7456)  
Bird,C.R., Smith,C.J., Ray,J.A., Moureau P., Bevan,M.W., Bird,A.S.,  
Hughes,S., Morris,P.C., Grierson,D. and Schuch,W.

TITLE The tomato polygalacturonase gene and ripening-specific expression  
in transgenic plants

JOURNAL Plant Mol. Biol. 11, 651-662 (1988)  
FEATURES  
source 1..7456  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"

/cultivar="Ailsa Craig"  
/db\_xref="taxon:4081"  
/clone\_1b="lambda EMBL3"  
repeat\_region 877..904  
/note="direct repeat 1"  
repeat\_region 908..935  
/note="direct repeat 1"  
TATA\_signal 1326..1335  
CAAT\_signal 1342..1347  
CAAT\_signal 1400..1403  
precursor\_RNA 1411..7244  
/note="primary transcript"  
exon 1411..1757  
/number=1  
TATA\_signal 1431..1435  
join(1479..1757,2416..2547,3327..3491,3696..3716,  
4260..4467,4567..4648,5602..5710,6139..6255,6788..7048)  
/codon\_start=1  
/product="polygalacturonase"  
/protein\_id="CA32235.1"  
/db\_xref="GI:295813"  
/db\_xref="GOA:P05117"  
/db\_xref="SWISS-PROT:P05117"  
/translation="MYIQNSILLITIFASISITGRSNVIDNLFKQYDNLBQEF  
ADPQAYLSYLSKNTLESNNIDKVDKNGIKIVNLVFGAKGDKTDNTAFQAMNBA  
CSSRTPVQFVPRKKNRYLKKQITFSGPCRSSISVKLFGLSLEASSKI SDYKRLMTAF  
DSVQNLVVGSGGTINNGOVWPPSSCKPKSKSPCRDAPALFPMNCKNLKVNLSKN  
AQQIHFKPESCINNVASNLIMINASKSPMDGVAVSNTOYIOISDTLIGDPCISIV  
SSQSNVQATNITCGRGHGISIGLSGSEAVSVNTVNAKTIIGAMGPRITWGG  
SSQASIKRLNVEMQDKYPIITIDNYTCRVECTIOQFSAVQKYNVYENIKTSATK  
VAIKPDCSTNPFCEGITMENIMLVGSSGKSEATCKNVHFNNAEHTPHOTSLEISD  
EALVINY"

intron 1758..2415  
/number=1  
exon 2416..2547  
/number=2  
intron 2548..3326  
/number=2  
exon 3327..3491  
/number=3  
intron 3492..3695  
/number=3  
exon 3696..3716  
/number=4  
intron 3717..4259  
/number=4  
exon 4260..4467  
/number=5  
intron 4468..4566  
/number=5  
exon 4567..4648  
/number=6  
intron 4649..5601  
/number=6  
exon 5602..5710  
/number=7  
intron 5711..6138  
/number=7  
exon 6139..6255  
/number=8  
intron 6256..6787  
/number=8  
exon 6788..7244  
/number=9  
polya\_site 7244

## ORIGIN

Query Match 28.2%; Score 462; DB 8; Length 7456;  
Best Local Similarity 100.0%; Pred. No. 5,3e-63;  
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1161 CAGTTTCAGCAGTTCAAGTGAATAATGTGTATGAGAAATCAAGGGCACAAGTGCA 1220  
|||

Db		6785	CAGTTTCAGAGGATTCAAGTGA AAAAATGTGGTGATGAGAAATATCAAGGCCACAAGTGCA	6844
Qy		1221	ACAAAGGTGGCCATAAATTTGATGTGCAGCA CAACCTTTCATGTGGAAGAATTATATG	1280
Db		6845	ACAAAGGTGGCCATAAATTTGATGTGCAGCA CAACCTTTCATGTGGAAGAATTATATG	6904
Qy		1281	GAGAAATATAATTTAGTAGGGGAAAAGTGA AATCCATCAGAGGCTACGTGCAAAAAATGTC	1340
Db		6905	GAGAAATATAATTTAGTAGGGGAAAAGTGA AATCCATCAGAGGCTACGTGCAAAAAATGTC	6964
Qy		1341	CATTTTAACAAATGCTGAACCATGTTACAC CACTGCACTTCTACTAGAAATTTCCAGAGAT	1400
Db		6965	CATTTTAACAAATGCTGAACCATGTTACAC CACTGCACTTCTACTAGAAATTTCCAGAGAT	7024
Qy		1401	GAACCTCTTTTGTAATTAATTAATTAATTA ATTAATTAATTAATTAATTAATTAATTAATTA	1460
Db		7025	GAACCTCTTTTGTAATTAATTAATTAATTA ATTAATTAATTAATTAATTAATTAATTAATTA	7084
Qy		1461	TATATCCAAATTAACAAATCTATATCTATG TATGAAATTAATTAATTAATTAATTAATTAAT	1520
Db		7085	TATATCCAAATTAACAAATCTATATCTATG TATGAAATTAATTAATTAATTAATTAATTAAT	7144
Qy		1521	TTGAAGTTTTAATAAGACTACTATGTAATTT CTATTTCTAGTCAAAAAGTTTGACAGATGT	1580
Db		7145	TTGAAGTTTTAATAAGACTACTATGTAATTT CTATTTCTAGTCAAAAAGTTTGACAGATGT	7204
Qy		1581	ACTTTTAATATGATCAAAAATATATAATG GTATTATTATATGA	1622
Db		7205	ACTTTTAATATGATCAAAAATATATAATG GTATTATTATATGA	7246
<hr/>				
RESULT_14				
LOCUS	AKL17942			
DEFINITION	Arabidopsis thaliana Ac3g57510 mRNA for putative endo-polygalacturonase, complete cds, clone: RAF19-13-H11.			
ACCESSION	AKL17942			
VERSION	AKL17942.1	GI:26450945		
KEYWORDS	FLI cDNA; CAP trapper.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;				
REFERENCE	1 Seki,M., Iida,K., Satou,M., Sakurai,T., Akiyama,K., Ishida,J., Nakajima,M., Enju,A., Kamiya,A., Natusaka,M., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.			
AUTHORS	Arabidopsis thaliana full-length cDNA Published Only in Database (2002) 2 (bases 1 to 1600)			
TITLE	Seki,M., Iida,K., Satou,M., Sakurai,T., Akiyama,K., Ishida,J., Nakajima,M., Enju,A., Kamiya,A., Natusaka,M., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.			
JOURNAL	Direct Submission Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:mseki@gsc.riken.go.jp, URL:http://pfweb.gsc.riken.go.jp, Tel:81-45-503-9625, Fax:81-45-503-9586)			
COMMENT	An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified lambdaBamBla-F1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site ( <a href="http://pfweb.gsc.riken.go.jp/">http://pfweb.gsc.riken.go.jp/</a> ) for further details.			
FEATURES	Location/Qualifiers			
SOURCE	1..1600 /organism="Arabidopsis thaliana" /mol_type="mRNA" /cultivar="Columbia"			

```

gene
CDS
    /db_xref="taxon:3702"
    /chromosome="3"
    /clone="RAFL19-13-H11"
    /note="common name: chale cress"
    1..1600
    /gene="At3g57510/T8H10_110"
    100..1395
    /gene="At3g57510/T8H10_110"
    /codon_start=1
    /product="putative endo-polygalacturonase"
    /protein_id="BAC42580.1"
    /db_xref="GI:26450946"
    /translation="MARCSRHIAVFLCVLMLSLCRALSNVDGDEGHEDGSPFSDSLKLNLDVSLISDEETLEASTYSVNFPAKGGKGTDDTQAFKAKKASCTGYTTFLVPGKTYILKSTRFRGPCSKLRNFILGLSTYSRSDYDKDMMHLLIEDVNNLSIDGSGTGIINGNGKWTWQNSCKIDKSKPCTAPATLLYNLKNLVKILRVKNAQI01SIEKCNLEENSVNETITAPGDSPTNDGJHITWTQNI RNSNSIDGDDCI SIEDGTQNIIFPLTCGPGHGIGSLGDDNSKAYSGIVADKAFESDNGVRIKTYOQSGSTARKIKRQNIEMENVKNPILIIDDQYCDKDKCEDQSAVQVKNVYKNIISGTSNDVATLNLGSEKTPCGGIVLENYKIKGTASCKNAVVKQGTVPKCS"

```

	Query Match	27.5%;	Score	449.2;	DB	8;	Length	1600;			
	Best Local Similarity	62.3%;	Pred.	40.7;	76-61;						
	Matches	788;	Conservative	0;	Mismatches	443;	Indels	33;	Gaps	4;	
Qy	277	TTAATGTA	CTTAA	CTTGAG	AGCTTA	AGGGTGA	TGGAAAA	CAATATG	ATATATG	CATTG	336
Db	302	TTAGTGT	TTCAAA	CTTCGAG	CCCAA	AGATG	AAAAA	ATGATG	ATCTACT	CAGGCTT	361
Qy	337	AGCAGCA	TGGAA	TGAAG	CACTT	CATCTA	GAACA	CACCTG	TTCAA	TTTGTG	366
Db	362	AGAAAGA	TGAA	GAAGCA	TGTTCA	CAATATG	AGTACT	CTTCTT	CTTGTTCT	TAAG	421
Qy	397	ACAAGAT	TTATCT	CTCAG	CAATCA	CTTTT	CAGGTC	CAATG	AGATCT	CTTAATTC	456
Db	422	GAAGAC	CTTATCT	CTTAA	GTCTA	CTGATTT	TGAGGCC	CAATG	AAATCT	CTTAGT	481
Qy	457	TAAAGAT	TTTTG	ATCC	CTTAG	AAGAC	CTATG	TAAAT	TTTCAG	CTACAA	516
Db	482	TTTGAT	CTTAA	GGCA	CTTTT	ATCAG	CACTAC	GAAG	CGTTG	ATTAACAA	541
Qy	517	TTTGAT	TGCTTT	TGATAG	TGTT	CAAAATTT	AGTTAG	TGAGAG	G-----	-AGAACTA	570
Db	542	ATTGGCT	TATCTT	TAGAG	AGCTT	TAACAATCT	ATCAATG	ACG	GGCTCG	ACGGAA	601
Qy	571	TCATGG	CAATG	GCAC	AGTATG	GTGGG	CAAGTTCT	TGCAAA	ATTAAT	TAATC	630
Db	602	TTAATG	CCAA	CGAAAA	CCCTG	GGGAG	AACTAT	GC	AAATATG	CAAACTTA	661
Qy	631	GCAGGA	TGCA	CCAC	CGCCTT	AACTCT	TGCAATTT	GCAAAA	ATTTGA	AGTGA	690
Db	662	GCA	CAAAA	AGCTCC	ACAGGCT	CTTATCTT	ATATCA	ATTTAA	GAATTTG	ATGTGA	721
Qy	691	TAAAG	ATAAAA	ATGCA	CAACA	ATTATAT	CAAAATTT	AGTCA	TGACTAT	GTGTAG	750
Db	722	TGAGG	TGAAAA	AAACCG	CAAGAT	TGCAATTT	CAATG	AAATG	CAACAA	ATTGA	781
Qy	751	CTTCAAA	TTTGAT	TATCA	TGCTT	CAGAA	AGACCC	AAATAC	TGATG	AGTCCAT	810
Db	782	TTAGTAA	TGTTG	AGATCA	CTGCT	CCGGG	CGATG	TCCCA	CA	CAATG	841
Qy	811	CAAA	TACTCA	TATAT	CAAA	TATCTG	TATCTAT	TGAA	CAAG	GTGAT	870
Db	842	CTAAT	ACTCA	AAAA	CAATTC	AGSTCT	CAAACT	CAGAT	ATCG	AAACAG	901
Qy	871	CAAT	TTTCT	TGATCT	CAAAAT	GTGA	GGCC	CAAA	TATTTA	CTTGTG	930
Db	902	CCAT	TGAG	TGAA	CGCAAA	ATCTT	CAAA	CTTTG	ATTTAA	CTTGCG	961
Qy	931	GTAT	AAATTT	TGGA	AGTTG	AGATCT	GAAATTC	GAAG	CGTTAT	GTCTAT	990
Db	962	GGAT	CA	CACTTT	GGAGCT	TGGGG	AGCA	CAATTC	GAAG	CTTATG	1021

QY 991 TAAATGAAGCAAAATTTATCGTGGCCGAAATGAGTTAGCATCAAGACTTGGCAGGGAG 1050  
|||  
Db 1022 TGGATGATGCTAAATTTCTCTGAGAGTGCATAGAGTTAGATTAGACTTATCAAGGGAG 1081  
|||  
QY 1051 GATCTGACAGCACTGACCAATCAAAATTTCTGAATGGGAAATGCAAGCGTTAACTATC 1110  
|||  
Db 1082 GATCAGCACTGACCAAGAACATTTAAATTTCAAAATATTTGAAATGAAACGTCACAAATC 1141  
|||  
QY 1111 CCATTAATTAATGACCAAAATCTATTGTGATGAGTTGACCATGATATATACAAGTTTTCAG 1170  
|||  
Db 1142 CCATCTAATATCGACCGAGACTACTGCGA---CAAGACAAATGCGGAAGACCAAGAGTCCG 1198  
|||  
QY 1171 CAGTTCAAGTGAATAATGTGCTGTATGAGAAATTCAGGGCACAAGTGCACAAAGGTGG 1230  
|||  
Db 1199 CAGTGCAGATGAAACGTTGTGTACAAAGACATCTGTGTACGAGCGCTACGAGATGTGG 1258  
|||  
QY 1231 CCATTAATTTGATTTGACAGCAAACTTCCATGTGAAAGAAATTTAAATGAGAAATATA 1290  
|||  
Db 1255 CCATTAACGTTGAATTTGACCGGAGAACTATCCATGTCAAGGAGTTGTGCTTGAGAACTGA 1318  
|||  
QY 1291 ATTTAGTGAAGGAAAGTGGAAAACCATCAGAGGCTACGTCAAAAATGTCATTTTAA 1350  
|||  
Db 1319 AATTAAGAGGAGAA-----CAGCTTCTTGCAAAAATGCCAATGTTAAA 1363  
|||  
QY 1351 ATGCTGAACATTTACACGACTGCACTTCACTTGAAATTTCAAGAGATGAAGCTCTT 1410  
|||  
Db 1354 ATCAAGGACCGCTTCTCTTAATGCTCTTAACGCTAATTAATGTAATGCA----- 1416  
|||  
QY 1411 TGTATATATTAATTTTACTATAGATCTTCATATATATGACAAATATATATATACAA 1470  
|||  
Db 1417 --CATACATATTTACATAGATGATGATGATATTTATATATATGATATATGATGATCACT 1474  
|||  
QY 1471 TAAACAAATCTATATCTATGATATGAAATATTTATTAATATGATGAGATGAAGTTT 1530  
|||  
Db 1475 GCATGCTCTTAAATACATGATGTAAGATATATGAAAAAATAGTCGAAAGTTTGTTTA 1534  
|||  
QY 1531 AATA 1534  
|||  
Db 1535 CAT 1538  
|||  
RESULT 15  
AX412563  
LOCUS AX412563 1296 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 327 From Patent WO0222675.  
ACCESSION AX412563  
VERSION AX412563.1 GI:21445021  
KEYWORDS  
SOURCE  
ORGANISM Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1  
REFERENCE  
AUTHORS Glazebrook, J., Wang, X., Dangl, J.L., Bulgem, T. and Zhu, T.  
TITLE Plant genes, the expression of which are altered by pathogen  
infection  
JOURNAL Patent: WO 0222675-A 327 21-MAR-2002;  
SYNGENTA PARTICIPATIONS AG (CH); UNIVERSITY OF NORTH CAROLINA AT  
CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Kun (US); Dangl,  
Jeffrey L. (US); Bulgem, Thomas (US)  
FEATURES  
Source Location/Qualifiers  
1..1296  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"  
ORIGIN  
Query Match 27.2%; Score 445.2; DB 6; Length 1296;  
Best Local Similarity 64.7%; Pred. No. 3,4e-60;  
Matches 720; Conservative 0; Mismatches 368; Indels 24; Gaps 3;

QY 277 TTAATGACTTAAGCTTTGAGGCTAAGGGTGAATGAAAAAATATATGATTAATTTGACTTTG 336  
|||  
Db 203 TTAGTGTTCCTCAACTTCGGAGCCAAAGAGATGAAAAATCATATGATATCTCAAGCCTTCA 262  
|||  
QY 337 AGCAGCATGGAATGAGCATGTTCACTTAGAACACCTGTTCAATTTGGTTCCTAAA 396  
|||  
Db 263 AGAAAGCATGGAAGAAAGCATGTTCAACAAATGAGTTACTTCTTGTTCTCTTAAAG 322  
|||  
QY 397 ACAAGATATATCTTCTCAGCAAAATCACTTTTCAGTCCATAGCAATCTCTCATTTTCAG 456  
|||  
Db 323 GAAAGCATATCTCTTAACTCTTACGATTTTAGAGGCCAATGCAAAATCCTTAACGTAAT 382  
|||  
QY 457 TAAAGATTTTGGATCTTGAAGCATCTAGTAAATTTTCAGACTCAAAAGATAGAAAGG 516  
|||  
Db 383 TTCAGATCTTAGGCACTTATAGCATCTAGCAAAAGTTTCAAAAGACAAAAC 442  
|||  
QY 517 TTTGATTTGCTTTTGAATAGTTCATAAATTTAGTTGTTGAGAGAGA-----GGAAGTA 570  
|||  
Db 443 ATTTGCTTAATCTTAAGAGACGTTAACATCTATCAATCGACGGTGGCTCGACGGGAATTA 502  
|||  
QY 571 TCAATGGCAATGGAACAATATGTTGGCCAGTTCTTGCAAAATTAATTAATCACTGCCAT 630  
|||  
Db 503 TTAATGGCAACGGAATAACCTGTGTGCGAGACTCATGCAAAATCGCAAAATCTTAACCAT 562  
|||  
QY 631 GCAGGAGTGCACCAACGGCTTAACTTCTGGAATTTGCAAAAATTTGAAAGTGAATATATC 690  
|||  
Db 563 GCACAAAAGCTCCACCGGCTCTTACTTATACAAATTAAGATTTGAATGGAACAAATC 622  
|||  
QY 691 TAAAGATTAATAATGCAACAATTCATATTAATTTGATGATGCACTAATGTTGTAG 750  
|||  
Db 623 TGAGGATGAATAATCGGACGAGATTCAGATTTCAATGGAATAATGCAAAAGTTGAAG 682  
|||  
QY 751 CTTCAATTTGATGATGATGCTTCAAGAAAGAGCCCAATCTGATGAGTGCACATGAT 810  
|||  
Db 683 TTAGTAATGTTGAATGATCTGCTCCGGGAGATGTCACAAACAAATGATATCCAAATGA 742  
|||  
QY 811 CAATATCTCAATATATCAAAATATCTGATATCTATTTATGGAACAGGTGATATGATTT 870  
|||  
Db 743 CTAATACTCAAAACATTTGAGATCTCCAACTCAGATATGGAACAGGTGATATGATTT 802  
|||  
QY 871 CAATTTGTTCTGATCTCAAAATGTGCAAGGCCACAATAATTAATCTTGTTGTCAGGTCA 930  
|||  
Db 803 CCATTTGAGATGGAACGCAAAATCTTCAAAATCTTGAATTAATCTTGCGGCCCGGTCAG 862  
|||  
QY 921 GTATTAATGATGGAAGCTTGAATCTGGAATTTCAAGAGCTTAATGCTTAATGTTACG 990  
|||  
Db 863 GGAATGCAATTTGGAGCTTGCGGACGCAATTTGAAAGCTTATGTCTCGGGAATTAATG 922  
|||  
QY 991 TAAATGAAGCCAAATTTATCGGTGCGCAAAATGAGTTAGATCAAGACTTGGCAGGGAG 1050  
|||  
Db 923 TGGATGTGCTAAGTTCTCTAGAGTGCATATGAGATTAGATTAAAGCTTATCAAGGGAG 982  
|||  
QY 1051 GATCTGCAAGCTTAGCAACATCAAAATTTCTGAATGTGAAAAATGCAAGACGTTAAGTATC 1110  
|||  
Db 983 GATCAGGACTGCAAGCAACATTAATTTCAAAATATTCGATGAGAAAAACGTCAAGAAATC 1042  
|||  
QY 1111 CCATTAATATAGACCAAAATCTATTGTGATGAGTTGAAACCTGATATACAAGTTTTCAG 1170  
|||  
Db 1043 CCATCATATATGACCGGACTACTGCGA---CAAGACAAATGCGAAGACCAAGAGTGG 1099  
|||  
QY 1171 CAGTTCAAGTGAATAATGTGTATGAGAAATATCAAGGGCACAAGTGCACAAAGAGTGG 1230  
|||  
Db 1100 CAGTGCAGATGAAAAAGTTGTGTAACAAGAACATATCTGATGAGAGCCCTAGAGATGTGG 1159  
|||  
QY 1231 CCATTAATTTGATTTGACGACCAATCTTCCATGTGAAAGATTTAATGAGAAATATA 1290  
|||  
Db 1160 CGATTAAGTTGAATTTGACGAGCAAGATATCCATGTCAAGGAGATGTGTGTTAGAAACGTGA 1219  
|||  
QY 1291 ATTTAGTGAAGGGAAGTGAAGAAACCATCAGAGGCTAGTGAAGAAATCTCATTTTACA 1350  
|||  
Db 1220 AATTAAGAGAGGAA-----CAGCTTCTTGCAAAAATGCAATGTTTAAA 1264  
|||  
QY 1351 ATGCTGAACATGTTACACCACTGCACTTCA 1382

Tue Jul 27 08:55:39 2004

us-10-018-604-1.rge

Page 16

Db 1265 ATCAAGGACCGTTTCTCTAAATGCTTAA 1296

Search completed: July 27, 2004, 01:30:44  
Job time : 6486 secs

Query Match	Score	DB 1;	Length
Best Local Similarity	Pred. No.	7.3e-194;	
100.0%;			
100.0%;			

Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60  
 XX 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60  
 Db 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60

QY 61 SNNNDKVDKNGIKVINVLSFGAKGDKTYDNIAPFQANNEACSSRTPVQFVVPKKNY 120  
 Db 61 SNNNDKVDKNGIKVINVLSFGAKGDKTYDNIAPFQANNEACSSRTPVQFVVPKKNY 120

QY 121 LKQITFSGPCRSSISVKIFGSLFASAKSIDYKDRRLMIAFDSYQNLVGGGGTINNGQV 180  
 Db 121 LKQITFSGPCRSSISVKIFGSLFASAKSIDYKDRRLMIAFDSYQNLVGGGGTINNGQV 180

QY 181 WPPSSCKINKSLPCRDAPALTFWNCNKLKVNLLKSKNAQOIHKEPSCNVAASMLMN 240  
 Db 181 WPPSSCKINKSLPCRDAPALTFWNCNKLKVNLLKSKNAQOIHKEPSCNVAASMLMN 240

QY 241 ASAKSPNTDGVHVSNTQYIQTIIIGTGDCCISIVSGSNVQATNITCGPHGISIGSL 300  
 Db 241 ASAKSPNTDGVHVSNTQYIQTIIIGTGDCCISIVSGSNVQATNITCGPHGISIGSL 300

QY 301 GSGNSEAVYNSVTVNEAKIIGAENGVRIKTWQGGSGQASNIKFLANEMODVKYPIIIDON 360  
 Db 301 GSGNSEAVYNSVTVNEAKIIGAENGVRIKTWQGGSGQASNIKFLANEMODVKYPIIIDON 360

QY 361 YCDRVEPCIQOFSAVQKVVYENIKGTSATKVAIKFDCSTNPPCGIIMENINLVGESG 420  
 Db 361 YCDRVEPCIQOFSAVQKVVYENIKGTSATKVAIKFDCSTNPPCGIIMENINLVGESG 420

QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNYY 457  
 Db 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNYY 457

RESULT 2  
 ID AAP94619 standard; protein; 457 AA.  
 AC AAP94619;  
 XX 25-MAR-2003 (revised)  
 DT 21-JUN-1990 (first entry)  
 XX Polygalacturonase (PG) cDNA gene product.  
 DE Polygalacturonase gene; tomato; ds.  
 KM Polygalacturonase gene; tomato; ds.  
 XX Lycopersicon esculentum.  
 OS US4801540-A.  
 XX 31-JAN-1989.  
 PD 02-JAN-1987; 87US-00000201.  
 PF 28-MAR-1986; 86US-00845676.  
 PR 17-OCT-1986; 86US-00920574.  
 XX (CALJ) CALGENE INC.  
 PA Hiatt WR, Sheehy RE, Shewmaker CK, Kridl JC, Knauf V;  
 PI WPI; 1989-053640/07.  
 DR N-PSDB; AAN91112.  
 XX Tomato polygalacturonase gene - used for modulating expression in plant  
 PT cells or directing expression of heterologous peptide(s).  
 XX Disclosure; Page; 8pp; English.  
 CC cDNA of gene may act as a probe to the genomic sequence, provides a means  
 of modulating the production of PG and acts as a source of the transit

CC peptide which may be joined to heterologous peptides directing them to  
 CC cell wall. (Updated on 25-MAR-2003 to correct PA field.)

QY 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60  
 XX 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60  
 Db 1 MWIORSILLLIIIFASSISTCRSNVIDNLFQVYDNLLEGEFAHDPQAYLSYSKNE 60

QY 61 SNNNDKVDKNGIKVINVLSFGAKGDKTYDNIAPFQANNEACSSRTPVQFVVPKKNY 120  
 Db 61 SNNNDKVDKNGIKVINVLSFGAKGDKTYDNIAPFQANNEACSSRTPVQFVVPKKNY 120

QY 121 LKQITFSGPCRSSISVKIFGSLFASAKSIDYKDRRLMIAFDSYQNLVGGGGTINNGQV 180  
 Db 121 LKQITFSGPCRSSISVKIFGSLFASAKSIDYKDRRLMIAFDSYQNLVGGGGTINNGQV 180

QY 181 WPPSSCKINKSLPCRDAPALTFWNCNKLKVNLLKSKNAQOIHKEPSCNVAASMLMN 240  
 Db 181 WPPSSCKINKSLPCRDAPALTFWNCNKLKVNLLKSKNAQOIHKEPSCNVAASMLMN 240

QY 241 ASAKSPNTDGVHVSNTQYIQTIIIGTGDCCISIVSGSNVQATNITCGPHGISIGSL 300  
 Db 241 ASAKSPNTDGVHVSNTQYIQTIIIGTGDCCISIVSGSNVQATNITCGPHGISIGSL 300

QY 301 GSGNSEAVYNSVTVNEAKIIGAENGVRIKTWQGGSGQASNIKFLANEMODVKYPIIIDON 360  
 Db 301 GSGNSEAVYNSVTVNEAKIIGAENGVRIKTWQGGSGQASNIKFLANEMODVKYPIIIDON 360

QY 361 YCDRVEPCIQOFSAVQKVVYENIKGTSATKVAIKFDCSTNPPCGIIMENINLVGESG 420  
 Db 361 YCDRVEPCIQOFSAVQKVVYENIKGTSATKVAIKFDCSTNPPCGIIMENINLVGESG 420

QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNYY 457  
 Db 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNYY 457

RESULT 3  
 ID AAR32107 standard; protein; 457 AA.  
 AC AAR32107;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 05-JUL-1993 (first entry)  
 XX Polygalacturonase.  
 DE Polygalacturonase.  
 KM polygalacturonase; PG; pectin esterase; PG; expression regulation;  
 KW fruit softening enzymes; flowering plants; fruiting plants;  
 KW antisense RNA.  
 XX Lycopersicon esculentum.  
 OS EP532060-A1  
 XX 17-MAR-1993.  
 PD 06-NOV-1987; 92EP-00117411.  
 PF 11-NOV-1986; 86GB-00026879.  
 PR (ICIL) IMPERIAL CHEM IND PLC.  
 PA (ZENE) ZENECA LTD.  
 XX Bridges IG, Grierson D, Schuch WW;  
 XX

DR WPI: 1993-087084/11.  
 DR N-PSDB; AAC8453.  
 XX  
 PT Recombinant DNA for flowering and fruiting plants e.g. tomatoes ripening  
 PT control - comprises base sequence for transcription contg. inverted  
 PT sequence of bases complementary to bases in anti sense ribonucleic acid  
 PT encoding softening enzymes, or gene expression regulation.  
 XX  
 XX Example 12; Fig 1; 20pp; English.  
 CC This is the sequence of polygalacturonase from clone pTOM6. The clone was  
 CC used to isolate the PG promoter in the construction of a vector encoding  
 CC antisense RNA to the PG cDNA and PG gene. This would be useful to  
 CC regulate the expression of the fruit softening enzymes in flowering and  
 CC fruiting plants. Such antisense RNA would delay fruit softening. (Updated  
 CC on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to correct  
 CC PF field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-  
 CC AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 457 AA;  
 Query Match 100.0%; Score 2390; DB 2; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-194;  
 Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WVIQNSILLILLIIFASSISTCRSNVIDNLFKQYDNLLEOFHADFQAYLSYLSKNIE 60  
 DB 1 WVIQNSILLILLIIFASSISTCRSNVIDNLFKQYDNLLEOFHADFQAYLSYLSKNIE 60  
 QY 61 SNNNDKVDKNGIKYINVLFGAKGDKTYNIAFEQAMNEACSSRTVPQFVPPKKNYL 120  
 DB 61 SNNNDKVDKNGIKYINVLFGAKGDKTYNIAFEQAMNEACSSRTVPQFVPPKKNYL 120  
 QY 121 LKQITFSGPCRSSISVKIFGSLBASISKIDYKDRRLMTAFPSVQNLVVGSGGTINGNGY 180  
 DB 121 LKQITFSGPCRSSISVKIFGSLBASISKIDYKDRRLMTAFPSVQNLVVGSGGTINGNGY 180  
 QY 181 WPPSSCKINKSLPCRDAPALTFFWNCCKLKVNNLKSKNAQOIHIKFEBCTNVVASNLMIN 240  
 DB 181 WPPSSCKINKSLPCRDAPALTFFWNCCKLKVNNLKSKNAQOIHIKFEBCTNVVASNLMIN 240  
 QY 241 ASAKSPNTDGVHVSNTQYIOISDTIIGDGCISIVSGSQNVQATNITCGPHGISTSL 300  
 DB 241 ASAKSPNTDGVHVSNTQYIOISDTIIGDGCISIVSGSQNVQATNITCGPHGISTSL 300  
 QY 301 GSGNSEAVSVNTVNEAKTIGAENGVRITKTOGSGGQASNIKFLNVEMQDYKYPITIDON 360  
 DB 301 GSGNSEAVSVNTVNEAKTIGAENGVRITKTOGSGGQASNIKFLNVEMQDYKYPITIDON 360  
 QY 361 YCDRVEPTIQFSAVQNVVYENIKGTSATKVAIKFPCSTNFPCEGIMENINIVGSEG 420  
 DB 361 YCDRVEPTIQFSAVQNVVYENIKGTSATKVAIKFPCSTNFPCEGIMENINIVGSEG 420  
 QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEIDEDALLNY 457  
 DB 421 KPSEATCKNVHFNNAEHVTPHCTSLSEIDEDALLNY 457  
 RESULT 4  
 AAB48338 standard; protein; 457 AA.  
 XX  
 AC AAB48338;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Tomato polygalacturonase (PG) enzyme.  
 XX  
 KM Pectin; pectin methyltransferase; PME; polygalacturonase; PG; tomato; pTOM6;  
 KM food product; yogurt; milk; fruit juice; whey drink; de-esterification.  
 XX  
 OS Lycopersicon esculentum.  
 XX

PN WO200078982-A1.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 15-JUN-2000; 2000WO-1B000865.  
 XX  
 PR 17-JUN-1999; 99GB-00014209.  
 XX  
 PA (DANI-) DANISCO AS.  
 XX  
 PI Christensen TME, Kreiberg JD;  
 XX  
 DR WPI: 2001-091573/10.  
 DR N-PSDB; AAC8453.  
 XX  
 PT Modifying pectin, for foodstuffs preparation, involves transforming host  
 PT having pectin methyltransferase (PME) and polygalacturonase (PG) activity by  
 PT silencing PG activity, to increase PME to PG ratio.  
 XX  
 PS Disclosure; Fig 1; 78pp; English.  
 CC The invention provides a new method for modifying pectin that involves  
 CC providing a host having pectin methyltransferase (PME) activity and  
 CC polygalacturonase (PG) activity, transforming the host by silencing PG  
 CC activity to provide an increased PME to PG ratio, preparing a PME extract  
 CC from the transformed host, and using the PME extract to modify pectin. A  
 CC PME modified pectin is useful for foodstuffs preparation, and to impart  
 CC an increased functionality to food products such as yogurt, milk/fruit  
 CC juice and whey drinks. PME is useful to reduce the number of ester groups  
 CC in a pectin in a block-wise manner, and to de-esterify two or more  
 CC adjacent galacturonic acid residue of a pectin on at least substantially  
 CC all of the pectin chains. The present sequence represents a PG enzyme  
 CC encoded by a pTOM6 cDNA  
 CC  
 SQ Sequence 457 AA;  
 Query Match 100.0%; Score 2390; DB 4; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-194;  
 Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WVIQNSILLILLIIFASSISTCRSNVIDNLFKQYDNLLEOFHADFQAYLSYLSKNIE 60  
 DB 1 WVIQNSILLILLIIFASSISTCRSNVIDNLFKQYDNLLEOFHADFQAYLSYLSKNIE 60  
 QY 61 SNNNDKVDKNGIKYINVLFGAKGDKTYNIAFEQAMNEACSSRTVPQFVPPKKNYL 120  
 DB 61 SNNNDKVDKNGIKYINVLFGAKGDKTYNIAFEQAMNEACSSRTVPQFVPPKKNYL 120  
 QY 121 LKQITFSGPCRSSISVKIFGSLBASISKIDYKDRRLMTAFPSVQNLVVGSGGTINGNGY 180  
 DB 121 LKQITFSGPCRSSISVKIFGSLBASISKIDYKDRRLMTAFPSVQNLVVGSGGTINGNGY 180  
 QY 181 WPPSSCKINKSLPCRDAPALTFFWNCCKLKVNNLKSKNAQOIHIKFEBCTNVVASNLMIN 240  
 DB 181 WPPSSCKINKSLPCRDAPALTFFWNCCKLKVNNLKSKNAQOIHIKFEBCTNVVASNLMIN 240  
 QY 241 ASAKSPNTDGVHVSNTQYIOISDTIIGDGCISIVSGSQNVQATNITCGPHGISTSL 300  
 DB 241 ASAKSPNTDGVHVSNTQYIOISDTIIGDGCISIVSGSQNVQATNITCGPHGISTSL 300  
 QY 301 GSGNSEAVSVNTVNEAKTIGAENGVRITKTOGSGGQASNIKFLNVEMQDYKYPITIDON 360  
 DB 301 GSGNSEAVSVNTVNEAKTIGAENGVRITKTOGSGGQASNIKFLNVEMQDYKYPITIDON 360  
 QY 361 YCDRVEPTIQFSAVQNVVYENIKGTSATKVAIKFPCSTNFPCEGIMENINIVGSEG 420  
 DB 361 YCDRVEPTIQFSAVQNVVYENIKGTSATKVAIKFPCSTNFPCEGIMENINIVGSEG 420  
 QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEIDEDALLNY 457  
 DB 421 KPSEATCKNVHFNNAEHVTPHCTSLSEIDEDALLNY 457

RESULT 5  
 ABR43936  
 ID ABR43936 standard; protein; 456 AA.  
 XX  
 AC ABR43936;  
 XX  
 DT 11-AUG-2003 (first entry)  
 XX  
 DE Tomato endopolygalacturonase enzyme.  
 XX  
 XX Plant; pectin transester synthase; xyloglucan; polylysine; heparin;  
 XX  
 XX anticoagulant; endopolygalacturonase; enzyme; tomato.  
 XX  
 OS Lycopersicon esculentum.  
 XX  
 PN WO2003017950-A2.  
 XX  
 PD 06-MAR-2003.  
 XX  
 PF 03-SEP-2002; 2002WO-US028066.  
 XX  
 PR 31-AUG-2001; 2001US-0316777P.  
 XX  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX  
 PI Albersheim P, Djeleneo-Albersheim I, Darvill A;  
 XX  
 DR WPI; 2003-421044/39.  
 XX  
 XX Formation of ester or amide bond between ester and alcohol or amine  
 PT involves treating the ester or its acid or salt with plant pectin  
 PT transester synthase in the presence of alcohol or amine.  
 XX  
 PS Disclosure; Fig 18; 67pp; English.  
 XX  
 XX The invention relates to forming an ester or amide bond between monomeric  
 CC or polymeric ester or its acid or salt and monomeric or polymeric alcohol  
 CC or amine by treating the ester or its acid or salt with a plant pectin  
 CC transester synthase in the presence of alcohol or amine under conditions  
 CC to form ester or amide bond. The method is useful for forming an ester or  
 CC amide bond between monomeric or polymeric ester or its acid or salt (e.g.  
 CC homogalacturonate) and monomeric or polymeric alcohol or amine. It is  
 CC useful for producing pectin-based polymers e.g., xyloglucan or D- or L-  
 CC polylysine, useful for the slow release of compounds in the body e.g.,  
 CC heparin with anticoagulant or other pharmaceutical properties. The  
 CC present sequence represents an endopolygalacturonase enzyme from tomato  
 XX  
 XX  
 SQ Sequence 456 AA;  
 Query Match 97.8%; Score 2338.5; DB 6; Length 456;  
 Best Local Similarity 98.9%; Pred. No. 1.7e-189;  
 Matches 452; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
 YY 1 MVIQRNSILLILIIIFASSISTCRSNVIDNLFKQVYDNIIEQFAHDFQAYLSYLSKNIE 60  
 YY 1 MVIQRNSILLILIIIFASSISTCRSNVIDNLFKQVYDNIIEQFAHDFQAYLSYLSKNIE 60  
 DB 1 MVIQRNSILLILIIIFASSISTCRSNVIDNLFKQVYDNIIEQFAHDFQAYLSYLSKNIE 60  
 YY 61 SNNNIDKVDKNGIKVINLVLSFGAKGDKTYDNTAFEQANNEACSSRTPVQFVVPKKNYLL 120  
 DB 61 SNNNIDKVDKNGIKVINLVLSFGAKGDKTYDNTAFEQANNEACSSRTPVQFVVPKKNYLL 120  
 YY 121 LKQITTEGPRCRSSISVYKIFGSLFASISIDYKDRRLMIAFDSVONLVVGGGTTINGNGOV 180  
 DB 121 LKQITTEGPRCRSSISVYKIFGSLFASISIDYKDRRLMIAFDSVONLVVGGGTTINGNGOV 180  
 YY 121 LKQITTEGPRCRSSISVYKIFGSLFASISIDYKDRRLMIAFDSVONLVVGGGTTINGNGOV 180  
 DB 121 LKQITTEGPRCRSSISVYKIFGSLFASISIDYKDRRLMIAFDSVONLVVGGGTTINGNGOV 180  
 YY 181 MWFSCKINKSLPCRDAPALTFMNCNKLKYNLKSNAQOIHKFESCTNVVASNLMIN 240  
 DB 181 MWFSCKINKSLPCRDAPALTFMNCNKLKYNLKSNAQOIHKFESCTNVVASNLMIN 240  
 YY 181 MWFSCKINKSLPCRDAPALTFMNCNKLKYNLKSNAQOIHKFESCTNVVASNLMIN 240  
 DB 181 MWFSCKINKSLPCRDAPALTFMNCNKLKYNLKSNAQOIHKFESCTNVVASNLMIN 240  
 YY 241 ASAXSEPTDGVHVSNTYIQTSDTIIGTGDPCISIVSGSQNVQNTITTCGHHGISIGSL 300  
 DB 241 ASAXSEPTDGVHVSNTYIQTSDTIIGTGDPCISIVSGSQNVQNTITTCGHHGISIGSL 300  
 YY 241 ASAXSEPTDGVHVSNTYIQTSDTIIGTGDPCISIVSGSQNVQNTITTCGHHGISIGSL 300  
 DB 241 ASAXSEPTDGVHVSNTYIQTSDTIIGTGDPCISIVSGSQNVQNTITTCGHHGISIGSL 300

YY 301 GSGNSEAVYVSNTVNEAKITIGAENGVRITWQGGSGQASNIKFLANEMQDVXPIIIDON 360  
 DB 301 GSGNSEAVYVSNTVNEAKITIGAENGVRITWQGGSGQASNIKFLANEMQDVXPIIIDON 360  
 YY 361 YCDRVEPCIQQPSANQVKNVVTENIKGTSATKAIKFDCTSTNFPFGGIIIMENINLYGSG 420  
 DB 361 YCDRVEPCIQQPSANQVKNVVTENIKGTSATKAIKFDCTSTNFPFGGIIIMENINLYGSG 420  
 YY 421 KPSEATCKNVHFNMAEHVTPHCTSLTSEDEALLVNY 457  
 DB 421 KPSEATCKNVHFNMAEHVTPHCTSLTSEDEALLVNY 456  
 RESULT 6  
 ABB92243  
 ID ABB92243 standard; protein; 438 AA.  
 XX  
 AC ABB92243;  
 XX  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE Herbicidally active polypeptide SEQ ID NO 1454.  
 XX  
 XX Herbicidal; plant; agriculture; herbicide.  
 XX  
 XX Arabidopsis thaliana.  
 XX  
 PN WO200210210-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP009892.  
 XX  
 PR 28-AUG-2001; 2001WO-EP009892.  
 XX  
 XX (FARB) BAYER AG.  
 XX  
 XX Tietjen K, Weidler M;  
 XX  
 DR WPI; 2002-269010/31.  
 XX  
 XX Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms.  
 XX  
 PS Claim 5; SEQ ID NO 1454; 261pp + Sequence Listing; English.  
 XX  
 XX The invention relates to identifying target proteins (ABB90790-ABB94016)  
 CC for herbicidally active compounds, comprising aligning and comparing  
 CC nucleic acid or amino acid sequences from plant with nucleic acid or  
 CC amino acid sequences from non-plant organisms using suitable search  
 CC parameters, where plant sequences having an E-value greater by a factor  
 CC of 3 than the E-value of most similar non-plant sequences are selected.  
 CC The polypeptides or nucleic acids encoding them are useful for  
 CC identifying modulators. The identified modulators are useful as  
 CC herbicides  
 XX  
 SQ Sequence 438 AA;  
 Query Match 49.6%; Score 1185; DB 5; Length 438;  
 Best Local Similarity 51.9%; Pred. No. 1.2e-91;  
 Matches 230; Conservative 76; Mismatches 121; Indels 16; Gaps 5;  
 YY 8 ILLIIIFASSISYCRSNV---IDNLFKQ---YYDNIIEQFAHDFQAYLSYLSKNIES 61  
 DB 5 ILLISVLLFTFLPSCFSSYFPENHRDDLPMSSNVTYETNRQHGHN-----TRNSHL 56  
 YY 62 NNNIDKVDKNGIKVINLVLSFGAKGDKTYDNTAFEQANNEACSSRTPVQFVVPKKNYLL 121  
 DB 57 KNRGIVAPRRSPRRFNNVTFGAKANGND-DSKAPMKMWEAACSSTGAIYIYAPKRDYML 115  
 YY 122 KQITTEGPRCRSSISVYKIFGSLFASISIDYKDRRLMIAFDSVONLVVGGGTTINGNGOV 181



```

Db      116 KAVTFSGPCSSLIITFYGRITKEMENPSDKERHMIWFENVNLRVEGGGRIDGNGHLM 175
QY      182 WPSCKINKSLPCRDAPALTFFMNCCKNLKVNNLKSKNAQOIHIKFESECTVVASNMINA 241
      176 WPSCKINPQLPLCLGAPLATVFECCNNLRVSNRLRNAQOMHLLTFQDCNKVKALNLMVMS 235
QY      242 SAKSEPTDGVHVSNTQYIQISDTITIGTDDCISIVSGSQNVQATNTTGGFHGISIGSLG 301
      236 PADSPNTDGIHVSGTGNIIIDISIVRTGDCISIVSGSEVRATIGTGGFHGISIGSLG 295
QY      302 SGNSEAYVSNVNTNNAKTIIGAENGVRITKMOGSSGQASINIKFLANVEMQVYPIITIDQNY 361
      296 EDNSEAYVSNVNVNKKATLIGTNGVRIKTWQGGHMAKNITIFQDIIMKNVTPIIINQDY 355
QY      362 CDRVEPCIOFSAVQVKNVYENIKGTSATKVAIKFPCSTNPFCEGIIENINLVGSEGG 421
      356 CRVEACPEQKSAVQVSNVLKXVIGTSSRPVAVKVCSSKNIPCRGISMQNVATLVDTQ 415
QY      422 P-SEATCKNVHFNNAEHVTPHCT 443
      416 DVSKASCSNVKLDTRGNVSLCT 438
Db

```

## RESULT 7

AA098178 standard; protein; 431 AA.

AA098178;

05-JUL-1999 (first entry)

Anther-specific ESJ2A protein.

ESJ2A gene; promoter; pollen; anther dehiscence; male sterile; transgenic plant.

Arabidopsis thaliana.

W09913089-A1.

18-MAR-1999.

11-SEP-1998; 98WO-GB002752.

11-SEP-1997; 97GB-00019359.

(BIOG-) BIOGEMMA UK LTD.

Roberts JA, Paul W, Craze M;

WPI; 1999-254279/21.

N-PSDB; AAX25011.

Generation of male sterile plants by controlling anther dehiscence.

Disclosure; Fig 3; 34pp; English.

This protein is encoded by the ESJ2A gene (see AAX25011) of Arabidopsis thaliana. The invention relates to the use of the ESJ2A promoter to reduce dehiscence and to create male sterile plants for use in hybrid seed production. The promoter is used to drive expression of a further nucleic acid sequence that results in prevention or reduction of anther dehiscence. For example, expression of the RNase barnase causes cell ablation, while expression of a plant hormone alters the developmental fate of a cell. Plants are produced that have phenotypically normal pollen grains, within phenotypically normal anthers, but in which the anthers do not dehisce and thus do not release the pollen grains. The system allows the female to be multiplied with the artificial male sterility gene in the homozygous state, since the female plant produces viable pollen. The system is suited to crops which have high seed multiplication, large amounts of pollen, and/or separate male and female inflorescences. These factors allow for easy collection of pollen from

CC non-dehiscing anthers, facile self-pollination and the minimisation of CC the area of plants that have to be self-pollinated manually. Such an CC ideal crop is the monocot maize, but the system is also applicable to CC e.g. wheat, barley, rice, fodder grass, banana, palm, orchid, tulip, CC lily, melon, cucumber, tomato, pepper and willow. It is also useful in CC the avoidance or reduction of pollen allergens and may be effective in CC the control of asthma caused by pollen release

XX Sequence 431 AA;

Query Match 47.2%; Score 1127; DB 2; Length 431;

Best Local Similarity 49.6%; Pred. No. 9.4e-87; Matches 223; Conservative 71; Mismatches 116; Indels 40; Gaps 6;

```

QY      7 SILLIIIPASSISTORNVID-----DNFKQVNDIIEQEFADHQAATLSYL 55
      9 APLFLCLMLSLCKALSSNVDDGYGHEDGSFESDSLKLANDDVL-----SLISSD 59
      56 SKNIESNNNIDKVDKNGIKVINVLSPGAKGDKTYDNIAPQAMNEACSSRTVPQFVVPK 115
      60 ETLLEAS-----TVSVNFGAKGDKTDDTQAFKKAMKACSTNGVTFILVPK 107
QY      116 NKNVYLKQITFSGPCRSSISVAKIFGSLSEASKISDYKDRRLMIAFDSVQNLVVGCG-GR 173
      108 GKTYLLKSTFRFGPCSKSLRNFIQLTSLASTKRSYDKDKRNHMLIEDVNNLSIDGGSTGI 167
QY      174 INNGQVWMPSSCKINKSLPCRDAPALTFFMNCCKNLKVNNLKSKNAQOIHIKFESECTNVV 233
      168 INNGKTMQNSCKIDKSKPCCKAPALTLYNLKUNVKNLRKNAQOIQISIEKCKVYE 227
QY      234 ASNLMINASAKSPNTDGVHVSNTQYIQISDTITIGTDDCISIVSGSQNVQATNTTGGPH 293
      228 VSNVEITAPDGSBNTDGIHITNTQNRVNSNDIGTDDCISIEDGTQNIPIPLTGGPH 287
QY      294 GISIGSLGSGNSEAYVSNVNTNNAKTIIGAENGVRITKMOGSSGQASINIKFLANVEMQVY 353
      288 GISIGSLGSDNSKAYVSGINVDGAKFSESNDGRIKTYOGSGSTANIKRQNRIMENVN 347
QY      354 PIITDQNYCDRVEPCIOFSAVQVKNVYENIKGTSATKVAIKFPCSTNPFCEGIIENINLV 413
      348 PIITDDYCDK-DKCEDQSAVQVKNVYKNSGISATVAILTNLSEKIPCGIIVENV 406
QY      414 NLVSGSKPSEATCKNVHFNNAEHVTPHCT 443
      407 KIKG-----GTASCKNANVKNQGVSPKCS 431
Db

```

## RESULT 8

AA041324 standard; protein; 431 AA.

AA041324;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 51400.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127452P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140685P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147312P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 11-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154038P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0156599P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.

```

PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159293P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```

Query Match 47.2%; Score 1127; DB 3; Length 431;  
 Best Local Similarity 49.6%; Pred. No. 9, 4e-87;  
 Matches 223; Conservative 71; Mismatches 116; Indels 40; Gaps 6;

```

QY 7 STILLIIIFASSISICRSNVID-----DNLFKQVYNIIIEQFAHDFQAYLSY 55
DB 9 AVFLCVLMLSLCKALSSNVDDGYGHEDGSFESDSILKLNDDVL-----SLISSD 59
QY 56 SKNIESNNNIDKVDKNGIKVINVLSPGAKGDKGTVDNIAFEQAMNEACSSRTTPVQFVVPK 115
DB 60 ETTLEAS-----TVSVSNFGAKGDKDTDTQAFKAKMKKACSTNGVTTFLVPK 107
QY 116 KKNYLLKQITFSGPCRSSISVKIFGSLBASISKISDYKRRRLWIAPDSVQNLVVG3--GT 173
DB 108 GKTYYLLKSTRFRGPKSLRNFIQILGTLASSTRSDYKDKNMHLLIEDVNNLSIDGSGTGI 167
QY 174 INGNQVWMPSSCKINKSLPCRDATPALTFFNCKLKNVNNLSKSKAQQOIHIFBESCTNVV 233
DB 168 INGNKRTWQNSCKIDKSKPCTKAPALTFLVNLKNLVNLRVKNAAQOIQISIECKNKYE 227
QY 234 ASNLMTNINSAKSPNTDGVHVSNTQYIOISDTIIGTGDICISIVSGSQNVQATNITCGPQH 293
DB 228 VSNVEITPAGSPNPDGHIHTNTQINRVNSNDIGTGDICISIEDGTQNLQITDLTGCGRH 287
QY 294 GISISLISGNSSEAVVSNVTYNEAKIIGAENGVRIKTWQSGSGQASNIKFLVEMQDVYK 353
DB 288 GISISLIDDDNSKAVVSGINVDGAKFSSDNGVRIKTYQSGSGTAKNIKFFQIRHENVKN 347
QY 354 PIIDQNTCDRVEPTIQOFSAYQVKNVYENIKGTSATKVALKFPDCSTNFPPEEGIMENI 413
DB 348 PIIDQDYCDK-DKCEDQSAVQVKNVYKNIISGTSATDVALTINCSEKYPQOGIVLENV 406
QY 414 NLVSGSKPSEATCKNHFNNNAHYTPCT 443
DB 407 KIRG-----GTASCKNANVKNQGTVPKCS 431

```

RESULT 9  
 ID ABB92704 standard; protein; 431 AA.  
 XX ABB92704  
 AC ABB92704;

```

XX 31-MAY-2002 (first entry)
DT Herbicidally active polypeptide SEQ ID NO 1915.
XX Herbicidally active polypeptide SEQ ID NO 1915.
DE Herbicidally active polypeptide SEQ ID NO 1915.
XX Herbicidally active polypeptide SEQ ID NO 1915.
XX Herbicidally active polypeptide SEQ ID NO 1915.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN WO200210210-A2.
XX WO200210210-A2.
PD 07-FEB-2002.
XX 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP009892.
XX 28-AUG-2001; 2001WO-EP009892.
XX 28-AUG-2001; 2001WO-EP009892.
XX (FARB ) BAYER AG.
XX Tietjen K, Weidler M;
XX MPI; 2002-269010/31.
XX MPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
XX organisms.
XX Claim 5; SEQ ID NO 1915; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
XX for herbicidally active compounds, comprising aligning and comparing
XX nucleic acid or amino acid sequences from plant with nucleic acid or
XX amino acid sequences from non-plant organisms using suitable search
XX parameters, where plant sequences having an E-value greater by a factor
XX of 3 than the E-value of most similar non-plant sequences are selected.
XX The polypeptides or nucleic acids encoding them are useful for
XX identifying modulators. The identified modulators are useful as
XX herbicides
XX
XX Sequence 431 AA:

```

Query Match 47.2%; Score 1127; DB 5; Length 431;  
 Best Local Similarity 49.6%; Pred. No. 9, 4e-87;  
 Matches 223; Conservative 71; Mismatches 116; Indels 40; Gaps 6;

```

QY 7 STILLIIIFASSISICRSNVID-----DNLFKQVYNIIIEQFAHDFQAYLSY 55
DB 9 AVFLCVLMLSLCKALSSNVDDGYGHEDGSFESDSILKLNDDVL-----SLISSD 59
QY 56 SKNIESNNNIDKVDKNGIKVINVLSPGAKGDKGTVDNIAFEQAMNEACSSRTTPVQFVVPK 115
DB 60 ETTLEAS-----TVSVSNFGAKGDKDTDTQAFKAKMKKACSTNGVTTFLVPK 107
QY 116 KKNYLLKQITFSGPCRSSISVKIFGSLBASISKISDYKRRRLWIAPDSVQNLVVG3--GT 173
DB 108 GKTYYLLKSTRFRGPKSLRNFIQILGTLASSTRSDYKDKNMHLLIEDVNNLSIDGSGTGI 167
QY 174 INGNQVWMPSSCKINKSLPCRDATPALTFFNCKLKNVNNLSKSKAQQOIHIFBESCTNVV 233
DB 168 INGNKRTWQNSCKIDKSKPCTKAPALTFLVNLKNLVNLRVKNAAQOIQISIECKNKYE 227
QY 234 ASNLMTNINSAKSPNTDGVHVSNTQYIOISDTIIGTGDICISIVSGSQNVQATNITCGPQH 293
DB 228 VSNVEITPAGSPNPDGHIHTNTQINRVNSNDIGTGDICISIEDGTQNLQITDLTGCGRH 287
QY 294 GISISLISGNSSEAVVSNVTYNEAKIIGAENGVRIKTWQSGSGQASNIKFLVEMQDVYK 353
DB 288 GISISLIDDDNSKAVVSGINVDGAKFSSDNGVRIKTYQSGSGTAKNIKFFQIRHENVKN 347
QY 354 PIIDQNTCDRVEPTIQOFSAYQVKNVYENIKGTSATKVALKFPDCSTNFPPEEGIMENI 413
DB 348 PIIDQDYCDK-DKCEDQSAVQVKNVYKNIISGTSATDVALTINCSEKYPQOGIVLENV 406

```

QY 414 NTVGSGKRESEATCKNVHNNNAEHTPHCT 443  
Db 407 KIRG----GTASCKANVNNQGTSPKCS 431  
RESULT 10  
AAG41323  
ID AAG41323 standard; protein; 463 AA.  
AC AAG41323;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51399.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131044P.  
XX 30-APR-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 07-MAY-1999; 99US-0132487P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 18-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136783P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140685P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 20-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0147038P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 06-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.

PR	13-AUG-1999;	99US-0146684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	30-AUG-1999;	99US-0151080P.
PR	31-AUG-1999;	99US-0151030P.
PR	01-SEP-1999;	99US-0151438P.
PR	07-SEP-1999;	99US-0151930P.
PR	10-SEP-1999;	99US-0152363P.
PR	13-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match	47.2%;	Score 1127;	DB 3;	Length 463;
Best Local Similarity	49.6%;	Pred. No. 1e-86;		
Matches 223;	Conservative 71;	Mismatches 116;	Indels 40;	Gaps 6
QY	7	SIIILIIIFASSISTCRSNVID-----DNLFKQYVDNIILQGFAPHDQAVSYL	55	
	:	:	:	:
	:	:	:	:
Db	41	AVFLCVLMLSLCKRLSNVDDGYGHEHDSFESDSLKKANNVDL-----SLISSD	91	
	:	:	:	:
QY	56	SKNIESNNIDKVKNGKIVINVLGFGAKGDKTYNIIAFEQAMNAGCSRRTPVQFVVPK	115	
	:	:	:	:
	:	:	:	:

Db	92	ETTLEAS-----TWSVSNFGAGDGDGDTDTQAFKAMKKAACSNVTTTELVPK	135
Qy	116	NKNYLLKQITPESGCRSSISVYKIFGSLSEASSKSIDYKDBRLMIAFDSVQNLVVGAG--GT	1737
Db	140	CKTYLLKSTRFRGPCCKLNFQILGLTSLASTKRDYKDKNHWLILEDVNNLSDGSGSTGI	199
Qy	174	INGNQVWMPSCCKINKSLPCSDAPALTALTFMNCNKLKNNLKSKNAQQHIKESCTNVV	233
Db	200	INGNCKTWMONSCKIDKSKPCTKAPALTALTYLNKLNINVLRYKNAQQIOISIEKCKVE	255
Qy	234	ASNIMINSAKSPNTDGVHVSNTQVYQIISDITLIGTGDDCISITVSGSQNVQATNITCGPGH	295
Db	260	VSNNVITLAPGDSFNPDGHITNTQIRVNSNDIGTDDCISIEDGTQNLQIFDITCGPGH	311
Qy	294	GISISLGSQNSEAVVNTVNEAKIIGAENGVRIKTWQSGSGQASNIKFLANEMQDPVKY	353
Db	320	GISISLGDNDNSKAVVSGINVDGAKFSESDNGRILKTYGGSGSTANIKFQNIEMENVXN	379
Qy	364	PIIIDQNYCDRVEPCIQPSAVQVKNVYENIKGTSATKYAIKFDGSTNFPCEGIMENI	411
Db	380	PIIIDQNYCDK-DKCEDQESAVQVKNVYKNIISGTSATDVAITLNGSEKYPGQIVLENV	433
Qy	414	NLVGESGKPEATCKNVHNNNAHVTPHCT	443
Db	439	KIKG-----GTASCKNANVKNQGTWSPKCS	463
RESULT 11			
AAG41325			
ID	AAG41325	standard, protein, 415 AA.	
XX	AC	AAG41325;	
XX	DT	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 51401.	
XX	KM	Protein identification; signal transduction pathway; metabolic pathway;	
KM	hybridisation assay; genetic mapping; gene expression control; promoter;		
KM	termination sequence.		
XX	OS	Arabidopsis thaliana.	
XX	PN	EP1033405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-00301439.	
XX	PR	25-FEB-1999; 99US-0121825P.	
PR	05-MAR-1999;	99US-0123180P.	
PR	09-MAR-1999;	99US-0123548P.	
PR	23-MAR-1999;	99US-0125788P.	
PR	25-MAR-1999;	99US-0126264P.	
PR	29-MAR-1999;	99US-0126785P.	
PR	01-APR-1999;	99US-0127462P.	
PR	06-APR-1999;	99US-0128234P.	
PR	08-APR-1999;	99US-0128714P.	
PR	16-APR-1999;	99US-0128645P.	
PR	19-APR-1999;	99US-0130077P.	
PR	21-APR-1999;	99US-0130449P.	
PR	23-APR-1999;	99US-0130510P.	
PR	23-APR-1999;	99US-0130891P.	
PR	28-APR-1999;	99US-0132044P.	
PR	30-APR-1999;	99US-0132407P.	
PR	04-MAY-1999;	99US-0132484P.	
PR	05-MAY-1999;	99US-0132485P.	
PR	06-MAY-1999;	99US-0132486P.	
PR	07-MAY-1999;	99US-0132487P.	
PR	11-MAY-1999;	99US-0132486P.	
PR	14-MAY-1999;	99US-0134218P.	

PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142802P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 04-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0147935P.  
PR 11-AUG-1999; 99US-0148171P.  
PR 12-AUG-1999; 99US-0148319P.  
PR 13-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157553P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 18-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.

```

PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 46.8%; Score 1119.5; DB 3; Length 415;
Best Local Similarity 50.6%; Pred. No. 3.8e-86;
Matches 223; Conservative 68; Mismatches 107; Indels 43; Gaps 7;

QY 19 ISNCR---SNVID-----DNLKQVYNIIEQEPHDFQAYLSLKNIESNNN 64
DB 2 LSLCKALSSVNDGYGHEGDSFESDSLKLKLNDDVL-----SLISSDETLLEAS-- 50
QY 65 IDKVDKNGIKVINVLSPGAKGDKTYDNIAPFQAMNEACSSRTPVQFVPEKNNKNTLLKQI 124
DB 51 -----TVSVSNPGAKGDKTDTQAFKAMKACSGTNGVTTFELVPEKGTLLKST 100
QY 125 TRSGPCRSSISVKTIFGSLASSKI SDYKRRRLIAPDSYONLVVGG--GTINGQVWV 182
DB 101 RFRGPKSLNFPQILGTLASASTKRSYKDKNHWLLEDVNNLSIDGSGTGIINGNGKTMW 160
QY 183 PSSCKINKSLPGDAPFALTFWCKNKLKNNLSKXNAQOIHIKESCTNVVASLMLNAS 242
DB 161 QNSCKIDKSKPCCTKAPALTLNKLNNLVKRNKAKQAQOIQISIECKNKVSVSWELTAP 220
QY 243 AKSPNTDGVHNSVNTQYIQTSDTITIGTDDCISIVSGSQNVQANTITGPGHGISIGSLG 302
DB 221 GDSPTNDGHIITQNRVNSNDIGTDDCISIEDGQNIQIPPLTGGPHGHSIGSLGD 280
QY 303 GNSGAYVSNVTNEAKITGABNGVRIKTWGGSGQASNIKFLVEMODVYKPIIIDQNYC 362
DB 281 DNGKAYVSGINLVGAKFSESDNGVRIKTYOGSGGTAKNIKFOINRMENVNKPIIIDDYC 340
QY 363 DREVPCTIQFSAVQVKNVYENIKGTSATKVAIKFDSTNPCEGIMENINLVGESKXP 422
DB 341 DK-DKEDQESAQVQVKNVYKNSIGTSATVAITLNCSEKIPCGIIVLENVKIKG----- 394
QY 423 SEATCKNHFNNAEHVTPHCT 443
DB 395 GTASCKNANVKNQGTVPKCS 415

RESULT 12
AB992075
ID ABB92075 standard; protein; 426 AA.
XX
AC ABB92075;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 1286.
XX
KM Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP009892.
XX
PR 28-AUG-2001; 2001WO-EP009892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M,

```

```

XX
XX WPI, 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX
XX Claim 5; SEQ ID NO 1286; 261pp + Sequence listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
XX Sequence 426 AA;
XX
QY 6 NSILLIIIFASSISTCRSNVIDNLFQVY-----DNILEQEPHDFQAYLSLKNIE 60
DB 6 NLTVFLMALLMFWSCKARISPNVYDHSYKFKSDSLIKRR--EDITGLRSPVRASLR 63
QY 61 SNNNIDKVDKNGIKVINVLSPGAKGDKTYDNIAPFQAMNEACSSRTPVQFVPEKNNKNTLL 120
DB 64 TPT-----TVSVSPFGAKGDKTDTQAFVAMKACSSNGAVMLVPEKGTLYL 112
QY 121 LKQITSGPCRSSISVKTIFGSLASSKI SDYKRRRLIAPDSYONLVVGG--GTINGNG 178
DB 113 LKSIQULGPNSLITVQIFETLSASQKRSYKDIKSWIMDGVNNLSVDGDTGVWGNG 172
QY 179 QVWPPSSCKINKSLPGDAPFALTFWCKNKLKNNLSKXNAQOIHIKESCTNVVASLML 238
DB 173 ETWQNSCKENKA-----KALTFVNSKSLIVKLVKRNAQOIQISIECKSNVQVSNV 225
QY 239 INASAKSPNTDGVHNSVNTQYIQTSDTITIGTDDCISIVSGSQNVQANTITGPGHGISIG 298
DB 226 VTAPADSPNTDGHITQNRVNSNDIGTDDCISIESSQNVQINDITCGPHGHSIG 285
QY 299 SLGSGNSEAYVSNVTNEAKITGABNGVRIKTWGGSGQASNIKFLVEMODVYKPIIID 358
DB 286 SLGDNSKAFVSGVYDGAKLSTGDNQVRIKTYOGSGGTASNIIFQNIQMDNVKNPPIIID 345
QY 359 QNYCDREVPCTIQFSAVQVKNVYENIKGTSATKVAIKFDSTNPCEGIMENINLVGE 418
DB 346 QDYCDK-SKCTTESNAQVKNVYRDISGSASNAITPFCNSKNYPCOGIIVLDRVNIKG- 403
QY 419 SGKPSKATCKNHFNNAEHVTPHCTS 444
DB 404 -----GKATCTNANVADKGAVALPQCNS 425

RESULT 13
AAW04268
ID AAW04268 standard; protein; 433 AA.
XX
AC AAW04268;
XX
DT 16-OCT-2003 (revised)
DT 14-DEC-1996 (first entry)
XX
DE Oilseed rape seed pod polygalacturonase.
XX
KM Polygalacturonase; seed pod; dehiscence; shatter; oilseed rape;
KM transgenic plant; antisense.
XX

```







***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 21:17:50 ; Search time 668 Seconds

(without alignment)  
10404.270 Million cell updates/sec

Title: US-10-018-604-1

Perfect score: 1636  
Sequence: 1 aatcttttcaatgacacag.....atatgaaaaaaaaaaaaa 1636

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	1636	2	AAQ38415
2	1636	100.0	1636	4	AAC84653
3	1635	99.3	1637	1	AAN80487
4	1611.6	98.5	1624	1	AAN91112
5	1042.4	63.7	5822	6	AAS20853
6	445.2	27.2	1296	7	ADA68531
7	445.2	27.2	1393	3	AAC46786
8	420.4	25.7	1657	2	AAT33994
9	420.4	25.7	1631	2	AAT63603
10	418.8	25.6	1657	2	AAZ22980
11	391.6	23.9	1280	7	ADA68538
12	375	22.9	1269	7	ADA70539
13	369	22.6	1182	7	ADA70540
14	331.8	20.3	2207	2	AAV27950
15	331.8	20.3	2207	2	AAV36973
16	331.8	20.3	2207	4	AAZ35696
17	331.8	20.3	2207	4	AAH47384
18	330.2	20.2	2207	1	AAN80748
19	308.4	18.9	1603	4	AAAD03721
20	296	18.1	1344	7	ADA70633
21	248.4	15.2	1524	4	AAI66167
22	248.4	15.2	1759	4	AAI66166
23	242.4	14.8	1772	2	AAT38521

24	236.6	14.5	1505	3	AAQ50895
25	228	13.9	1410	2	AAQ66054
26	228	13.9	1479	2	AAQ66051
27	228	13.9	1726	2	AAQ66054
28	226.4	13.8	1545	2	AAQ90156
29	225.4	13.8	1395	2	AAQ66050
30	224.8	13.7	1542	2	AAQ84045
31	224.8	13.7	1733	2	AAQ84044
32	224.8	13.7	1733	2	AAQ84048
33	221.4	13.5	1236	7	ADA69348
34	220.2	13.5	1380	2	AAQ84044
35	212	13.0	219	1	AAQ71212
36	212	13.0	219	2	AAQ14561
37	212	13.0	219	2	AAQ99158
38	208.6	12.8	1059	3	AAQ36613
39	177.2	10.8	1492	3	AAQ32994
40	175	10.7	442	6	ABU94109
41	175	10.7	442	7	ABX61023
42	174.8	10.7	1496	2	AAQ56767
43	171.2	10.5	1673	6	AAQ32961
44	151.8	9.3	1185	6	ABZ14715
45	149	9.1	1736	3	AAQ42340

#### ALIGNMENTS

##### RESULT 1

AAQ38415  
ID AAQ38415 standard; DNA; 1636 BP.

AC AAQ38415;

XX  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 05-JUL-1993 (first entry)  
XX  
DE Polylacturonase cDNA clone pTOM6.

KW polylacturonase; PG, pectin esterase; PE, expression regulation;  
KW fruit softening enzymes; flowering plants; fruiting plants;  
KW antisense RNA.

OS Lycopersicon esculentum.

XX  
FH Key Location/Qualifiers  
FT CDS S1..1422  
FT /tag= a  
FT /label= polylacturonase (PG)

XX  
EP532060-A1.

XX  
PD 17-MAR-1993.

XX  
PF 06-NOV-1987; 92EP-00117411.

XX  
PR 11-NOV-1986; 86GB-00026879.

XX  
PA (ICIL) IMPERIAL CHEM IND PLC.

XX  
PA (ZENE) ZENECA LTD.

XX  
PI Bridges IG, Grierson D, Schuch WW;

XX  
DR WPI: 1993-087084/11.

XX  
DR P-PSDB: AAR32107.

PT Recombinant DNA for flowering and fruiting plants e.g. tomatoes ripening  
PT control - comprises base sequence for transcription contg. inverted  
PT sequence of bases complementary to bases in anti sense ribonucleic acid  
PT encoding softening enzymes, or gene expression regulation.

XX  
PS Example 12; Fig 1; 20p; English.





QY 1501 TTAATATTAATGACGATGATTTAATAGACTACTATGATTTCTAATTTCTA 1560  
DB 1501 TTAATATTAATGACGATGATTTAATAGACTACTATGATTTCTAATTTCTA 1560  
QY 1561 GTCAAAAGTTGACGATGATTTAATAGACTACTATGATTTCTAATTTCTA 1620  
DB 1561 GTCAAAAGTTGACGATGATTTAATAGACTACTATGATTTCTAATTTCTA 1620  
QY 1621 GAAAAAAGTTGATTTAATAGACTACTATGATTTCTAATTTCTA 1636  
DB 1621 GAAAAAAGTTGATTTAATAGACTACTATGATTTCTAATTTCTA 1636  
RESULT 3  
AAN80487  
ID AAN80487 standard; DNA; 1637 BP.  
AC AAN80487;  
XX  
XX 25-MAR-2003 (revised)  
DT 20-NOV-1990 (first entry)  
XX  
XX Plasmid clone pTom6.  
DE  
XX Fruit ripening; polygalacturonase; pectin esterase; ss.  
KM  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FH 51..1424  
FT /\*tag= a  
FT CDS /label= polygalacturonase  
FT  
XX  
XX EP271988-A.  
XX  
XX PD 22-JUN-1988.  
XX  
XX PF 06-NOV-1987; 87EP-00309853.  
XX  
XX PR 11-NOV-1986; 86GB-00026879.  
XX  
XX PA (ICIL ) IMPERIAL CHEM IND PLC.  
XX (ZENE ) ZENECA LTD.  
XX  
XX PI Bridges IG, Schuch WW, Grierson D;  
XX  
XX DR MPI: 1988-169271/25.  
XX P-PSDB; AAP80299.  
XX  
XX  
XX Recombinant DNA comprising promoter and terminator sequences - useful in  
PT plants for altering ripening properties esp. in tomatoes.  
XX  
XX PS Disclosure; Page ?; 22pp; English.  
XX  
XX This cDNA clone, encoding polygalacturonase (PG), is used to produce  
CC antisense mRNA (with an inverted sequence to that of PG mRNA) which is  
CC inserted into a vector used to transform plants which thereafter have  
CC altered ripening properties. The inverted sequence and the PG mRNA form a  
CC double-stranded structure which inhibits expression of the PG mRNA. See  
CC also AAN80488. (Updated on 25-MAR-2003 to correct PA field.)  
XX  
XX SQ Sequence 1637 BP; 598 A; 232 C; 290 G; 517 T; 0 U; 0 Other;  
QY Query Match 99.3%; Score 1625; DB 1; Length 1637;  
DB Best Local Similarity 99.9%; Pred. No. 2,1e-275;  
Matches 1636; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 AATCTTTTCAATAGACAAAGTTTAAAAACCATACATATACATATATGATGTTATCC 60  
DB 1 AATCTTTTCAATAGACAAAGTTTAAAAACCATACATATATATATATATGATGTTATCC 60  
QY 61 AAGGAATAGATATTCCTTCCTCATTAATATTTTGGCTTATCAATTTCACTGTAGAA 120  
DB 61 AAGGAATAGATATTCCTTCCTCATTAATATTTTGGCTTATCAATTTCACTGTAGAA 120

DB 61 AAGGAATAGATATTCCTTCCTCATTAATATTTTGGCTTATCAATTTCACTGTAGAA 120  
QY 121 GCAATGTTATGATGACAAATTTATTCACAAAGTTTATGATAATTTCTTGAACAGAAAT 180  
DB 121 GCAATGTTATGATGACAAATTTATTCACAAAGTTTATGATAATTTCTTGAACAGAAAT 180  
QY 181 TTGCTCATGATTTTCAAGCTTATCTTCTATATTTGACCAAAATATGAAAGCAACATA 240  
DB 181 TTGCTCATGATTTTCAAGCTTATCTTCTATATTTGACCAAAATATGAAAGCAACATA 240  
QY 241 ATATATGCAAGGTGATTAATAAATGGGATTAAGGATTAATGATGATGATGATGATGAT 300  
DB 241 ATATATGCAAGGTGATTAATAAATGGGATTAAGGATTAATGATGATGATGATGATGAT 300  
QY 301 AAGGTGATGAAAAACATATGATTAATTTGATTTGACAAAGCATGATGATGATGATGAT 360  
DB 301 AAGGTGATGAAAAACATATGATTAATTTGATTTGACAAAGCATGATGATGATGATGAT 360  
QY 361 CATCTAGAACACTGTTTCAATTTTGTGTTCTTAAACAAAGATTTCTTCAAGCAAA 420  
DB 361 CATCTAGAACACTGTTTCAATTTTGTGTTCTTAAACAAAGATTTCTTCAAGCAAA 420  
QY 421 TCACTTTTCAAGTTCATGACGATCTTCTATTTCACTAAAGATTTTGGATCTTGAAG 480  
DB 421 TCACTTTTCAAGTTCATGACGATCTTCTATTTCACTAAAGATTTTGGATCTTGAAG 480  
QY 481 CATCTAGTAAATTTTCAAGTTCATGACGATCTTCTATTTCACTAAAGATTTTGGATCTT 540  
DB 481 CATCTAGTAAATTTTCAAGTTCATGACGATCTTCTATTTCACTAAAGATTTTGGATCTT 540  
QY 541 AAAATTTAGTTTGGAGAGAGAGAACTATCAATGGAATGGAATGGAATGGAATGGAAT 600  
DB 541 AAAATTTAGTTTGGAGAGAGAGAACTATCAATGGAATGGAATGGAATGGAATGGAAT 600  
QY 601 GTTCTTGGCAAAATTAATTAATCACTGCCATGAGGATGACAAAGGCTTTTAACTTCT 660  
DB 601 GTTCTTGGCAAAATTAATTAATCACTGCCATGAGGATGACAAAGGCTTTTAACTTCT 660  
QY 661 GGAATTTGCAAAATTTTGAAGTGAATTAATCAAGATTAATCAAGATTAATCAAGATTA 720  
DB 661 GGAATTTGCAAAATTTTGAAGTGAATTAATCAAGATTAATCAAGATTAATCAAGATTA 720  
QY 721 TCAATTTGAGTCATGACATATGTTGATGCTTCAATTTGATGATGATGATGATGATG 780  
DB 721 TCAATTTGAGTCATGACATATGTTGATGCTTCAATTTGATGATGATGATGATGATG 780  
QY 781 AGAGCCCAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
DB 781 AGAGCCCAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
QY 841 CTATTTATGGAACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
DB 841 CTATTTATGGAACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
QY 901 CCAAAATTTTACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
DB 901 CCAAAATTTTACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
QY 961 ATTCAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
DB 961 ATTCAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
QY 1021 ATGAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
DB 1021 ATGAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
QY 1081 TGAATGGAATGCAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATG 1140  
DB 1081 TGAATGGAATGCAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATG 1140  
QY 1141 GATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
DB 1141 GATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200



```

DB 904 TATTACTTGGGTCAGAGTCATGTAATGATTTGGAAAGCTTAGAGATCTGGAATTTGAGA 963
QY 968 AGCTTATGTGCTAATGTTA CTGTAATGAAAGCCAAATATCGGTCCGAAAAATGAGAT 1027
DB 964 AGCTTATGTGCTAATGTTA CTGTAATGAAAGCCAAATATCGGTCCGAAAAATGAGAT 1023
QY 1028 TAGGATCAAGACTTGGCAGGAGAGATCTGCAAGCATAGCAACATCTGAAATTTGAAATGT 1087
DB 1024 TAGGATCAAGACTTGGCAGGAGAGATCTGCAAGCATAGCAACATCTGAAATTTGAAATGT 1083
QY 1088 GGAATGCAAGAGCTTAGTATCCCATATATATAGACCAAAATCTGATGCAAGTTGA 1147
DB 1084 GGAATGCAAGAGCTTAGTATCCCATATATATAGACCAAAATCTGATGCAAGTTGA 1143
QY 1148 ACCATGTATACAAAGTTTTCAGCAGTTCAAGTGAAGTGTGTATGAAATATCAA 1207
DB 1144 ACCATGTATACAAAGTTTTCAGCAGTTCAAGTGAAGTGTGTATGAAATATCAA 1203
QY 1208 GGGCACAAGTGCACAAAGGTGGCCATTAATTTGATTGGCAGCAAACTTTCATGTGA 1267
DB 1204 GGGCACAAGTGCACAAAGGTGGCCATTAATTTGATTGGCAGCAAACTTTCATGTGA 1263
QY 1268 AGGAATTTATATGAGATATATTAATTTAGTAGGGGAAAGTGGAAAACCATCAGAGGCTAC 1327
DB 1264 AGGAATTTATATGAGATATATTAATTTAGTAGGGGAAAGTGGAAAACCATCAGAGGCTAC 1323
QY 1328 GTGCAAAAATGTCCATTTTAA CAATGTGAAACATGTTCACCAACCTGCACTTCACTTGA 1387
DB 1324 GTGCAAAAATGTCCATTTTAA CAATGTGAAACATGTTCACCAACCTGCACTTCACTTGA 1383
QY 1388 AATTTCAGAGATGAAGCTCTTTTGTATATATTTATTTATTTATTTATTTATTTATTTAT 1447
DB 1384 AATTTCAGAGATGAAGCTCTTTTGTATATATTTATTTATTTATTTATTTATTTATTTAT 1443
QY 1448 ATAGCAGATATGATATATATATATATATATATATATATATATATATATATATATATAT 1507
DB 1444 ATAGCAGATATGATATATATATATATATATATATATATATATATATATATATATATAT 1503
QY 1508 TAAATATGACGATGGAAGTTTAAATTAAGACTACTATGATTTCTATTTCTATGTCMAAA 1567
DB 1504 TAAATATGACGATGGAAGTTTAAATTAAGACTACTATGATTTCTATTTCTATGTCMAAA 1563
QY 1568 GTTTGACGATGCTATTTTATATGTAACAAATATATTAATGATTTATATGAAA 1625
DB 1564 GTTTGACGATGCTATTTTATATGTAACAAATATATTAATGATTTATATGAAA 1621

```

```

XX 14-AUG-2001; 2001MO-US025538.
PF 15-AUG-2000; 2000US-0225508P.
PR 07-AUG-2001; 2001US-00924197.
XX (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX Guterson N, Oeller P;
XX WPI; 2002-257599/30.
XX
XX Reducing the expression of a target gene in a cell, comprises expressing
XX in the cell an expression cassette comprising a promoter operably linked
XX to a sense or antisense targeting sequence and an inverted repeat of a
XX nopaline synthase gene.
XX
XX Claim 51; Page 37-39; 39pp; English.
XX
XX The present invention relates to an improved method for inhibiting the
XX expression of a target gene in a cell, by expressing in the cell an
XX expression cassette comprising a promoter operably linked to a sense or
XX antisense targeting sequence having substantial identity to a subsequence
XX of the target gene, and an inverted repeat (IR) of a subsequence of a NOS
XX (nopaline synthase) gene, where the IR is unrelated to the targeting
XX sequence. The expression cassette, pFP-IRN1 is constructed using a
XX figwort mosaic virus (FMV) promoter in which the 5'-untranslated leader
XX (5'-UTR) is derived from a plant heat shock 70 (hsp70) gene, the tomato
XX polygalacturonase (PG) gene and an IR of the terminator of the
XX Agrobacterium tumefaciens NOS gene. The method is useful for regulating
XX expression of endogenous plant phenotypes such as disease resistance,
XX expression of endogenous plant phenotypes such as disease resistance,
XX flavour, protein or nutritional characteristics. The improved gene
XX silencing construct is used in functional genomics to determine the
XX effect of regulating gene expression of a selected endogenous gene or
XX transgene. The method is simple and rapid, and is suitable for high-
XX throughput studies. Multiple transgenic constructs all containing the
XX same repeat element can be silenced at the same time, since the initial
XX silencing trigger mediated through the inverted repeat region will apply
XX to all of the transcripts. The present sequence represents the expression
XX cassette, pFP-IRN1
XX
XX Sequence 5822 BP; 1689 A; 1275 C; 1292 G; 1566 T; 0 U; 0 Other;
XX
XX Query Match 63.7%; Score 1042.4; DB 6; Length 5822;
XX Best Local Similarity 99.4%; Pred. No. 2.1e-173;
XX Matches 1046; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```



Db 1739 GCACCTAATGTTGACCTCAATTTGATGATCAATGCTTCAGCAAGGCCCAATCTG 1798  
Qy 796 ATGAGTCCATGATCAATTAATCTCAATATATCTGATCTATTTATTTGGAAG 855  
Db 1799 ATGAGTCCATGATCAATTAATCTCAATATATCTGATCTATTTATTTGGAAG 1858  
Qy 856 GTGATGATGATTTCAATGTTGTTGATCTCAAAATGTCAGGCCCAATATTTACTT 915  
Db 1859 GTGATGATGATTTCAATGTTGTTGATCTCAAAATGTCAGGCCCAATATTTACTT 1918  
Qy 916 GTGTCAGGTCATGATTAATTAATTTGAGGCTTGAATCTGGAATTCAGAGCTTGT 975  
Db 1919 GTGTCAGGTCATGATTAATTAATTTGAGGCTTGAATCTGGAATTCAGAGCTTGT 1978  
Qy 976 TGTCTAATGTTACTGTAATGATGACCAAAATTTGTCGCGAAATGAGATTGAGATCA 1035  
Db 1979 TGTCTAATGTTACTGTAATGATGACCAAAATTTGTCGCGAAATGAGATTGAGATCA 2038  
Qy 1036 AGACTTGGCAGGAGGATCTGGACAGATGACCAATTTCTGATGTTGGAATGC 1095  
Db 2039 AGACTTGGCAGGAGGATCTGGACAGATGACCAATTTCTGATGTTGGAATGC 2098  
Qy 1096 AAGACGTTAAGTATCCCATATATTTAGACCAAACTATTGATGTCGATGACCATGTA 1155  
Db 2099 AAGACGTTAAGTATCCCATATATTTAGACCAAACTATTGATGTCGATGACCATGTA 2158  
Qy 1156 TACACAGCTTTTCAGCAGTTCAAGTGAATAATGTCGTGATGAGATATCAAGGACAA 1215  
Db 2159 TACACAGCTTTTCAGCAGTTCAAGTGAATAATGTCGTGATGAGATATCAAGGACAA 2218  
Qy 1216 GTGCACAAAAGGTCGATTAATTTGATGTCGACCAAACTTTCCATGTCGAAGAAATTA 1275  
Db 2219 GTGCACAAAAGGTCGATTAATTTGATGTCGACCAAACTTTCCATGTCGAAGAAATTA 2278  
Qy 1276 TAAATGGAATATTAATTTAGTAGGGAAGTGAATAACATCAGAGGCTACGTCGAAA 1335  
Db 2279 TAAATGGAATATTAATTTAGTAGGGAAGTGAATAACATCAGAGGCTACGTCGAAA 2338  
Qy 1336 ATGTCCATTTTAAATGTCGAAATGTCGACCATGTCAGCTGCTACTGAAATTTTCA 1395  
Db 2339 ATGTCCATTTTAAATGTCGAAATGTCGACCATGTCAGCTGCTACTGAAATTTTCA 2398  
Qy 1396 AGGATGAAGCTCTTTGATTAATTTAATTTT 1427  
Db 2399 AGGATGAAGCTCTTTGATTAATTTAATTTT 2430

RESULT 6  
ADA68531  
ID ADA68531 standard; DNA; 1296 bp.  
XX  
AC ADA68531;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Arabidopsis thaliana gene, SEQ ID 557.  
XX  
KM Plant; bacterial infection; fungal infection; viral infection; ds.  
OS Arabidopsis thaliana.  
XX  
PN MO2003000898-A1.  
XX  
PD 03-JAN-2003.  
XX  
PF 22-JUN-2001; 2001MO-IB001105.  
XX  
PR 22-JUN-2001; 2001MO-IB001105.  
PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Kategiri F, Quan S, Tao Y, Whitlam S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
DR  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
PS Claim 6; SEQ ID NO 557; 899bp; English.  
XX  
CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.

Sequence 1296 bp; 424 A; 244 C; 293 G; 335 T; 0 U; 0 Other;  
Query Match 27.2%; Score 445.2; DB 7; Length 1296;  
Best Local Similarity 64.7%; Pred. No. 6.5e-69;  
Matches 720; Conservative 0; Mismatches 368; Indels 24; Gaps 3;

Qy 277 TTAATGTAATTAATGTTGAGTGAAGGTCGATGGAATAATGATTAATGATTTGATTTG 336  
Db 203 TTAATGTTTCAAACTTCGAGCCAAAGAGATGGAATAATGATTAATGATTTGATTTG 262  
Qy 337 AGCAGAGGATGATGAGATGATTTTCAATCTGATGACCACTGTTCAATTTGTCGTTTCA 396  
Db 263 AGAAGATGGAAGAAAGATGATGATCAAAATGAGATTAATGATTTGTCGTTTCA 322  
Qy 397 ACAAGATTAATCTTCTCAAGCAATCACTTTTCAGGTCGATGATGATTTGATTTGATTTG 456  
Db 323 GAAAGATTAATCTTCTCAAGCAATCACTTTTCAGGTCGATGATGATTTGATTTGATTTG 382  
Qy 457 TAAAGATTTTGGATCTTGAAGCAATCTGATGATTTTGAAGCAATCAAGATGAGAGC 516  
Db 383 TTAAGATCTTGAAGCAATCTTGAAGCAATCTGATGATTTTGAAGCAATCAAGATGAGAGC 442  
Qy 517 TTTGATGCTTTTGAATGATGATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 570  
Db 443 ATTTGCTTATCTTGAAGCAATCTTGAATTTTGAATTTTGAATTTTGAATTTTGA 502  
Qy 571 TCAATGCAATGAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 630  
Db 503 TTAATGCAAGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 562  
Qy 631 GCAAGGATGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 690  
Db 563 GCAAGGATGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 622  
Qy 691 TAAAGATTTTGAAGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGA 750  
Db 623 TGAAGGATTTTGAAGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGA 682  
Qy 751 CTTCATTAATTTGATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 810  
Db 683 TTAATGATTTGATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 742  
Qy 811 CAATTAATTTGATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 870  
Db 743 CTAAATCTCAAAACATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 802  
Qy 871 CAATTTGTTGATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 930  
Db 803 CCAATGAGATGAGCAAGCAAGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGA 862  
Qy 931 GTATTAATTTGAGATGATGATGTCGCAAGTTCTTGAATTTTGAATTTTGAATTTTGA 990

```
Db      | 863 | GAGTCGACATTGGAGGCTTGGGGACGACAACTTGGAAAGCTTATGCTCGGAAATTATG 922
Qy      | 991 | TAAATGAAGCCAAATTTATGCTGGCCGAAAATGAGTTAGATCAGACTTTGGCAGGAG 1050
Db      | 923 | TGGATGGTGTAAAGTTCTTGAGAGTGAACATGAGTTAGATTAAGACTTATCAGGGAG 982
Qy      | 1051 | GATCTGGACAAAGCTAGCAACATCAAATTTCTGATGATGCGAAATGCGAAGTTAGATC 1110
Db      | 983 | GATCAGGAACCTGCCAGAACACATTAAATTTCAAAATATTCGATGGAAGAGCTCAAGATC 1042
Qy      | 1111 | CCATTAATTTATAGCCAAACATATGCTGATGAGTGGATGCAATATATCAACAGTTTTCAG 1170
Db      | 1043 | CGATCATTAATCGACACGACTACTGCGCA----CAAGCAAAATGCGAAGACCAAGTCGG 1099
Qy      | 1171 | CAGTTCAAGTGAATAATGTGGTGTATGATGAATTCAGAGGCGCAAGTGCACAAAGTGG 1230
Db      | 1100 | CAGTGAAGTGAATAAAGCTGTGTATCAAGAACATATCTGTACAGAGCGCTACGATGTGG 1159
Qy      | 1231 | CCATTAATTTGATTTGACGACCAAACTTCCATGATGGAAGATATATATGAGATATTA 1290
Db      | 1160 | CGATTAACGTTGAATTTGACGACGAGATATCATGTCAAGGATTTGTCTTGAAGACGTGA 1219
Qy      | 1291 | ATTTAGTAGGGGAAGTGAATAACCATCAGAGGCTACGTGCAAAAATGTCCATTTTACA 1350
Db      | 1220 | AATTAAGAGGAGAA-----CAGCTTCTTGCAAAAATGCCAATGTTAA 1264
Qy      | 1351 | ATGCTGAACATGTTACACCACTGCACCTTCA 1382
Db      | 1265 | ATCAAGGACCGTTTCTCTTAATGCTCTTAA 1296
```

## RESULT 7

AAC46786 standard; DNA; 1393 BP.

ID AAC46786 standard; DNA; 1393 BP.

AAC46786;

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 51398.

Hybridisation assay; genetic mapping; gene expression control;  
protein identification; signal transduction pathway; metabolic pathway;  
promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

```
XX      | 25-FEB-1999; 99US-0121825P.
PR      | 99US-0123180P.
XX      | 05-MAR-1999; 99US-0123548P.
PR      | 09-MAR-1999; 99US-0125788P.
XX      | 23-MAR-1999; 99US-0126264P.
PR      | 25-MAR-1999; 99US-0126785P.
XX      | 29-MAR-1999; 99US-0127462P.
PR      | 01-APR-1999; 99US-0128234P.
XX      | 06-APR-1999; 99US-0128714P.
PR      | 08-APR-1999; 99US-0129845P.
XX      | 16-APR-1999; 99US-0130449P.
PR      | 19-APR-1999; 99US-0130777P.
XX      | 21-APR-1999; 99US-0130510P.
PR      | 23-APR-1999; 99US-0130891P.
XX      | 28-APR-1999; 99US-0131449P.
PR      | 30-APR-1999; 99US-0132048P.
XX      | 04-MAY-1999; 99US-0132484P.
PR      | 05-MAY-1999; 99US-0132485P.
```

```
PR      | 06-MAY-1999; 99US-0132486P.
PR      | 06-MAY-1999; 99US-0132487P.
PR      | 07-MAY-1999; 99US-0132863P.
PR      | 11-MAY-1999; 99US-0134256P.
PR      | 14-MAY-1999; 99US-0134218P.
PR      | 14-MAY-1999; 99US-0134219P.
PR      | 14-MAY-1999; 99US-0134221P.
PR      | 14-MAY-1999; 99US-0134370P.
PR      | 18-MAY-1999; 99US-0134768P.
PR      | 19-MAY-1999; 99US-0134941P.
PR      | 20-MAY-1999; 99US-0135124P.
PR      | 21-MAY-1999; 99US-0135133P.
PR      | 24-MAY-1999; 99US-0135629P.
PR      | 25-MAY-1999; 99US-0136021P.
PR      | 27-MAY-1999; 99US-0136392P.
PR      | 28-MAY-1999; 99US-0136782P.
PR      | 01-JUN-1999; 99US-0137222P.
PR      | 03-JUN-1999; 99US-0137528P.
PR      | 04-JUN-1999; 99US-0137502P.
PR      | 07-JUN-1999; 99US-0137724P.
PR      | 08-JUN-1999; 99US-0138094P.
PR      | 10-JUN-1999; 99US-0138540P.
PR      | 10-JUN-1999; 99US-0138847P.
PR      | 14-JUN-1999; 99US-0139119P.
PR      | 16-JUN-1999; 99US-0139452P.
PR      | 16-JUN-1999; 99US-0139453P.
PR      | 17-JUN-1999; 99US-0139492P.
PR      | 18-JUN-1999; 99US-0139454P.
PR      | 18-JUN-1999; 99US-0139455P.
PR      | 18-JUN-1999; 99US-0139456P.
PR      | 18-JUN-1999; 99US-0139457P.
PR      | 18-JUN-1999; 99US-0139458P.
PR      | 18-JUN-1999; 99US-0139459P.
PR      | 18-JUN-1999; 99US-0139460P.
PR      | 18-JUN-1999; 99US-0139461P.
PR      | 18-JUN-1999; 99US-0139462P.
PR      | 18-JUN-1999; 99US-0139463P.
PR      | 18-JUN-1999; 99US-0139464P.
PR      | 18-JUN-1999; 99US-0139750P.
PR      | 18-JUN-1999; 99US-0139763P.
PR      | 21-JUN-1999; 99US-0139817P.
PR      | 22-JUN-1999; 99US-0139899P.
PR      | 23-JUN-1999; 99US-0140353P.
PR      | 23-JUN-1999; 99US-0140354P.
PR      | 24-JUN-1999; 99US-0140695P.
PR      | 28-JUN-1999; 99US-0140823P.
PR      | 29-JUN-1999; 99US-0140991P.
PR      | 30-JUN-1999; 99US-0141287P.
PR      | 01-JUL-1999; 99US-0141842P.
PR      | 01-JUL-1999; 99US-0142154P.
PR      | 02-JUL-1999; 99US-0142055P.
PR      | 06-JUL-1999; 99US-0142390P.
PR      | 08-JUL-1999; 99US-0142803P.
PR      | 09-JUL-1999; 99US-0142920P.
PR      | 12-JUL-1999; 99US-0142977P.
PR      | 13-JUL-1999; 99US-0143542P.
PR      | 14-JUL-1999; 99US-0143624P.
PR      | 15-JUL-1999; 99US-0144005P.
PR      | 16-JUL-1999; 99US-0144085P.
PR      | 16-JUL-1999; 99US-0144086P.
PR      | 16-JUL-1999; 99US-0144325P.
PR      | 19-JUL-1999; 99US-0144331P.
PR      | 19-JUL-1999; 99US-0144332P.
PR      | 19-JUL-1999; 99US-0144333P.
PR      | 19-JUL-1999; 99US-0144334P.
PR      | 19-JUL-1999; 99US-0144335P.
PR      | 20-JUL-1999; 99US-0144332P.
PR      | 20-JUL-1999; 99US-0144684P.
PR      | 20-JUL-1999; 99US-0144814P.
PR      | 21-JUL-1999; 99US-0145086P.
PR      | 21-JUL-1999; 99US-0145088P.
PR      | 22-JUL-1999; 99US-0145085P.
PR      | 22-JUL-1999; 99US-0145087P.
```

Query	Best Local Similarity	27.2 %	Score 445.2	DR 3	Length 1393
Pr	21-OCT-1999	99US-0160814P			
Pr	21-OCT-1999	99US-0160815P			
Pr	22-OCT-1999	99US-0160980P			
Pr	22-OCT-1999	99US-0160981P			
Pr	22-OCT-1999	99US-0160989P			
Pr	25-OCT-1999	99US-0161404P			
Pr	25-OCT-1999	99US-0161405P			
Pr	25-OCT-1999	99US-0161406P			
Pr	26-OCT-1999	99US-0161359P			
Pr	26-OCT-1999	99US-0161360P			
Pr	26-OCT-1999	99US-0161361P			
Pr	28-OCT-1999	99US-0161920P			
Pr	28-OCT-1999	99US-0161922P			
Pr	28-OCT-1999	99US-0161932P			
Pr	29-OCT-1999	99US-0162142P			
Query Match	27.2 %	Score 445.2	DR 3	Length 1393	
Best Local Similarity	64.7 %	Pred. No. 6,6e-69			
Matches 720	Conservative 0	Mismatches 368	Indels 24	Gaps 3	
Pr	21-OCT-1999	99US-0160814P			
Pr	21-OCT-1999	99US-0160815P			
Pr	22-OCT-1999	99US-0160980P			
Pr	22-OCT-1999	99US-0160981P			
Pr	22-OCT-1999	99US-0160989P			
Pr	25-OCT-1999	99US-0161404P			
Pr	25-OCT-1999	99US-0161405P			
Pr	25-OCT-1999	99US-0161406P			
Pr	26-OCT-1999	99US-0161359P			
Pr	26-OCT-1999	99US-0161360P			
Pr	26-OCT-1999	99US-0161361P			
Pr	28-OCT-1999	99US-0161920P			
Pr	28-OCT-1999	99US-0161922P			
Pr	28-OCT-1999	99US-0161932P			
Pr	29-OCT-1999	99US-0162142P			
Query	277	TTAATGTAATCTTGGATCTTGGAGCTAAGGGTATGAGAAAACATATGATTAATATGCACTTGG	336		
Pr	300	TTAGTGTTCACAACTTCGAGCCCAAGAGATGGAAAACTGATGTACTCAGGCCCTTCA	359		
Query	337	AGAAAGCATGAAATGAAAGCATGTTCACTTGAACACCTGTTCATTTGGTTCCTAA	396		
Pr	360	AGAAAGCATGAAAGAAAGCATGTTCAACAAATGAGATTACTACTTTCTTGGTTCCTAAAG	419		
Query	397	ACAAAGATATATCTTCGACAAATACACTTTTCAGTCTCAATGCAATCTTCAATTCAG	456		
Pr	420	GAAAGCTTATCTCTTAAAGTCTACTGATTTAGAGGCCCAATGCAATCTTACATCT	479		
Query	457	TAAAGATTTTGGATCCTTAGAAGCATCTAGTAAATTTCAAGACTACAAAGATAGAGGC	516		
Pr	480	TTCAAGTCTCAGCACTTTATCAGCATCTACAAAGCTTCAATTTCAAAAGCAAAAACC	539		
Query	517	TTTGGATTCCTTGAATAGTGTTCAAAATTTAGTTTGGAGAGAGA-----GGAACTA	570		
Pr	540	ATTGCTTATCTTAAAGAGCGTTAAACAATCTATCAATGAGCGTGGCTCGAGGGAATTA	599		
Query	571	TCAATGGCAATGGAACAATATGCTGCCAAGTTCTTGCAAAATTAATATCACTGCCAT	630		
Pr	600	TTAATGGCAACGAAAAAATCTGGTGGCAAGTCAATGCAAAATGCAAAATCTAAGCCAT	659		
Query	631	GCAAGGATGACCAACGCGCTTAACCTCTGGAATTTGCAAAATTTGAAGATTAATC	690		
Pr	660	GCACAAAAGCTCCACGCGCTCTTACTTTATCAATTTAAAGAAATTTGAATGGAAGATC	719		
Query	691	TAAAGATTAATAATGACAAACAAATTCATCAATTTGAGTCATCTATGTTGATG	750		
Pr	720	TGAGGATGAATAATCGCAGCAGATTCAGATTTCAATTGAAATGCAACAAAGTTGAAG	779		
Query	751	CTTCAATTTGATGATCATGCTTCAGCAAAAGGCCCAATATCTGATGAGATCCATGAT	810		
Pr	780	TTAGTAATGTTGAGATCACTGCTCGGCGAGTAGTCCCAACAGATGATCCATATCA	839		
Query	811	CAATATCTCAATATTTCAATATCTGATCTATTTATTTGGAACGAGTATGATTTAT	870		
Pr	840	CTAATATCTCAAAACATTTGAGTCTTCACACTCAGATATCGAAGCGATGATATGATAT	899		
Query	871	CAATGTTTTCGATCTTCAAAATGNGACGCCAACAATATTAATCTGTCAGGCTCA	930		
Pr	900	CCATTTGAGGATGGAACGCAAAATCTTCAAACTTTGATTTAACTTGGGCCCGCGGTACG	959		
Query	931	GTAATGATTTGGAAGCTTAGAGATCGAAATTCAGAAAGCTTATGATGCTAATGTTACTG	990		
Pr	960	GGAATGAGATTTGGAAGCTTGGGGGAGCACAATTCGAAGCTTATGCTCGGGAATTAATG	1019		
Query	991	TAAATGAGCCCAAAATTTATGGGTCCGAAAAATGAGTTAGATCAAGATCTGGACAGGAG	1050		



Db 1307 TGGCATATGTTTATGACGTGGAATATCCATGCCAGATTTGTGCTTGAGATG 1366  
Qy 1288 TAAATTAGTAGGGGAAAGTGAAGAAACATCAGAGGCTTACGTAATGTCATTTTA 1347  
Db 1367 TGAACATTAAGAGGAA-----AAGCTTTCGCAAAATGTCATGTTA 1411  
Qy 1348 ACAATGCTGACATCTTACCCACCTGACCTTCACTGAAATTTGAGAGATGAAGCTC 1407  
Db 1412 AGGATTAAGGACCGCTTCTCTTAATGCCCTTAATTAAGTTGATTAATATATAC 1471  
Qy 1408 TTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1447  
Db 1472 ATAAATACGATTAATGCTGTTATAGATGCAATCTAAT 1511

RESULT 9  
AAT63603  
ID AAT63603 standard; cDNA to mRNA; 1631 BP.  
XX  
AC AAT63603;  
XX  
DT 17-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 08-JUL-1997 (first entry)  
XX  
DE Dehiscence zone-selective endopolysaccharuronase cDNA.  
XX  
KM Dehiscence zone; endo-polygalacturonase; seed shattering; oilseed rape;  
XX transgenic plant; antisense; ribozyme; ds.  
XX  
OS Brassica napus; cv. Topaz.  
XX  
FH Key  
FT CDS  
FT 95..1396  
FT /\*tag= a  
FT sig\_peptide  
FT 95..163  
FT /\*tag= b  
FT mat\_peptide  
FT 164..1393  
FT /\*tag= c  
FT primer\_bind  
FT complement(821..837)  
FT /\*tag= d  
FT primer\_bind  
FT complement(884..900)  
FT /\*tag= e  
FT primer\_bind  
FT 1059..1073  
FT /\*tag= f  
FT primer\_bind  
FT 1229..1245  
FT /\*tag= g  
FT misc\_difference  
FT 1439  
FT /\*tag= h  
FT /note= "base 1439 is given as n in the specification"

MO9713865-A1.  
PD 17-APR-1997.  
PF 04-OCT-1996; 96WO-EP004313.  
PR 06-OCT-1995; 95EP-00402241.  
PR 08-DEC-1995; 95EP-00203328.  
XX  
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
PI Ulskov P, Child R, Van Onckelen H, Prinsen E, Borkhardt B,  
XX Sander J, Peteren M, Burdard Poulsen G, Bolterman U;  
DR WPI; 1997-235901/21.  
XX  
FT Transgenic plant containing dehiscence zone selective chimeric gene - has

PT modified dehiscence properties, especially delayed pod dehiscence.  
XX  
PS Claim 6; Page 48; 65pp; English.  
CC  
CC cDNA clone X (AAT63603) includes a coding sequence for oilseed rape  
CC dehiscence zone (DZ)-selective endopolysaccharuronase (PG). DZ cDNA was  
CC subjected to PCR amplification using primers (see also AAT63605-08) based  
CC on conserved regions of PG amino acid sequences. PG-related clones (see  
CC also AAT63609-14) were identified, of which only Ip935-8 was specific to  
CC the DZ. This clone was used to screen a DZ-selective cDNA library prepd.  
CC from poly+A+ RNA isolated from the DZ 6 wk after anthesis, yielding clone  
CC X. Ip935-8 was also used to screen a genomic library to identify the DZ-  
CC selective PG gene promoter region (AAT63604). These sequences can be  
CC utilised in novel chimeric genes to modify the dehiscence properties of  
CC transgenic plants, partic. the pod dehiscence properties of B. napus.  
CC This can improve the seed yield from the transformed plants. (Updated on  
CC 25-MAR-2003 to correct PI field.) (Updated on 17-OCT-2003 to standardise  
CC OS field)  
XX  
SQ Sequence 1631 BP; 554 A; 291 C; 339 G; 446 T; 0 U; 1 Other;

Query Match 25.7%; Score 420.2; DB 2; Length 1631;  
Best Local Similarity 63.6%; Pred. No. 1.6e-64;  
Matches 717; Conservative 0; Mismatches 383; Indels 27; Gaps 4;

Qy 277 TTAATGACTTACGCTTGGAGCTTAAGGCTGATGGAAGAAACATGATTAATTTGATTTG 336  
Db 300 TTAGTGTTCGAATCTTCGAGCCGCAAGAGATGGAAGAAACGATTAATCTGAGCTTCA 359  
Qy 337 AGCAAGCATGGAATGAGACATGTCATCTAGAACACCTGTTCAATTTGGTTCCTAA 396  
Db 360 AGAAGCATGGAAGAGGATGTTCAACAAATGAGTACTTCTTTATTTCTTAAG 419  
Qy 397 ACAAGATTAATCTTCACCAAAATCACTTTGAGTTCATGATGATTTTATTTTAC 456  
Db 420 GAAAGCTTATCTCCTTAAGTCTATTAAGTTCAGAGGCCATCAATCTTACGATGCT 479  
Qy 457 TAAAGATTTTGGATCTTGAAGACATCTAGTAATTTTCAAGCTC---AAGATGAA 513  
Db 480 TCCAGATCTAGGACATTTTACGCTTTTACAAAGATGATGATTCAGTAATGACAGA 539  
Qy 514 GGGTTGATGCTTTTGTATGATGTTCAAAATTTA-----GTTGTGAGAGAGAGAA 567  
Db 540 ACCACTGGCTTATTTTGAAGAGCTTAATTAATTAATCAATGACGCGGCGGGGA 599  
Qy 568 CTATCAATGSCAATGACAAAGTATGTTGCCAAATTTTGAATTAATTAATCACTGC 627  
Db 600 TTGTGATGCAAGCAAGAAATATCTGTGCAAAACATGCAAAATCGCAAAATCTAAGC 659  
Qy 628 CATGACGGATGACCAACGAGGCTTAACCTTGGAAATGCAAAATTTGAAGTAAAT 687  
Db 660 CATGACAAAGCGGCAACGCTCTTACTCTTACCACTTAAGAAATTTGAATGTAAGA 719  
Qy 688 ATCTAAGAGTAAATGACAAACAAATTCATATCAAAATTTGAGTATGATGATGTTG 747  
Db 720 ATCTGAGAGTGAAGATGACAGAGATTTCAATTTGAAATGCAACAAATGTTG 779  
Qy 748 TAGCTTCAATTTGATGATCAATGCTTACGCAAGAGCCCAATATCTGATGATGATG 807  
Db 780 GCGTTAAGAAATGTTAAGATCACTGCTCTGCGAATGATGCCAAGCGATGATTCATA 839  
Qy 808 TATCAATATCTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 867  
Db 840 TCGTGTCTACTTAACAAACATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAAT 899  
Qy 868 TTTCAATTTCTTGTGATCTCAAAATGACAGGCAAAATTTTCTGTTGATGATGATG 927  
Db 900 TATCATTTGAGAGATGACGAAATGTTCAATTAATTAATTAATTTGATGATGATGATG 959  
Qy 928 ATGATTAAGTATGGAAGCTTATGATGATGATGATGATGATGATGATGATGATGATG 987  
Db 960 ATGATTAAGTATGGAAGCTTATGATGATGATGATGATGATGATGATGATGATGATG 1019

QY 988 CTGTAATGAAGCCAAATTTATGCTGCCGAAATGGAGTTAGATCAAGACTTGGCAGG 1047  
DB 1020 ATGTGATGTGCTAGCGCTCTCTGAGACTGACATGAGTAAGATCAAGCTTACGAGG 1079  
QY 1048 GAGGATCTGAGCAAGCAGTCAACATCAATTTCTGATGTGAAATGCAAGCGTTAAGT 1107  
DB 1080 GAGGGTCAGGAACTGCTAGAACTTTAATTCGAAAATTCGTATGATGATCAAGA 1139  
QY 1108 ATCCCATTAATTAAGCAAACTATTGATGAGTTGAAACATGATACACAGTTT 1167  
DB 1140 ATCCGATCATATGACAGCACTACTGCGA---CAAGGCAAAATGCGAACAGCAGAAAT 1196  
QY 1168 CAGAGATTCAGTAAATAATGTGTATAGAAATCAAGGCAAGTGCACAAAG 1227  
DB 1197 CTGGGTTCAAGTAAACATGCTGTATCTGAAACATTAAGGTCACAGCGCAACAGTG 1256  
QY 1228 TGCGCATTAATTTGATTTGACGACAAACTTTCCATGTGAGGAATTAATGAGAAATA 1287  
DB 1257 TGCGGATTAATTTAATTTGAGTGTGAAATTCATGCAAGGTATGTGCTGAGAAATG 1316  
QY 1288 TAAATTTAGTAGGGGAAAGTGGAAACCAATCAGAGGCTACGTGCAAAAATGTCCATT 1347  
DB 1317 TGAACATCAAAAGAGGAA-----AAGCTTTCTGGGAAATGTCAATGT 1361  
QY 1348 ACAATGCTGAACATGTTACACACACACTGCACTGACTAGAAATTTCA 1394  
DB 1362 AGGATTAAGGCACTGTTCTCTTAATGCTTAATTAATTAAGCTGA 1408

RESULT 10  
AAZ22980  
ID AAZ22980 standard; DNA; 1657 BP.  
XX AAZ22980;  
XX AC  
XX 10-JAN-2000 (first entry)  
XX DT  
XX DE Brassica napus Sac66 protein encoding DNA.  
XX DE  
XX KW Signal transduction protein; dehiscence; male sterile plant; D22 gene;  
XX KW shatter resistance; oilseed rape; Sac66 protein; ss.  
XX OS Brassica napus.  
XX PN WO9949046-A1.  
XX PD 30-SEP-1999.  
XX PF 22-MAR-1999; 99WO-GB000905.  
XX PR 20-MAR-1998; 98GB-00006113.  
XX PA (BIOG-) BIOGENMA UK LTD.  
XX PI Wyatt P, Roberts JA, Whitelaw C;  
XX DR WPI; 1999-580449/49.  
XX DR P-PSDB; AAY42649.  
XX PT A nucleic acid encoding a signal transduction protein involved in plant  
XX PS dehiscence, useful for producing shatter resistant male sterile plants.  
XX PS Example 6; Fig 15; 71pp; English.  
CC The invention provides a nucleic acid encoding a signal transduction  
CC protein involved in the process of dehiscence. The nucleic acids and  
CC proteins are useful for regulating or controlling dehiscence of a pod or  
CC an anther in a plant, useful in the production of shatter resistance or  
CC methods, etc. may be used in the production of shatter resistance or  
CC shatter-delayed plants such as oilseed rape (Brassica napus). The present  
CC sequence represents the DNA encoding a B. napus Sac66 protein  
SQ Sequence 1657 BP; 541 A; 313 C; 352 G; 451 T; 0 U; 0 Other;

Query Match 25.6%; Score 418.8; DB 2; Length 1657;  
Best Local Similarity 62.4%; Pred. No. 2,8e-64;  
Matches 736; Conservative 0; Mismatches 417; Indels 27; Gaps 4;  
QY 277 TTAATGACTTACTAGCTTTGAGGCTTAAGGTGATGAGAAACATATGATATATGATTCATT 336  
DB 350 TTAATGTTGAACTTGGAGCAAAAGGTATGAGAAACCGATGATCTCAGGCTTTCA 409  
QY 337 AGCAAGCATGAAATGAAAGCATGTTCACTAGAAACACCTGTTCAATTTGTGTTCTTAAAA 396  
DB 410 AGAAGCATGAAAGAGGCACTGTTCAACAAATGAGAGTACTACTTTGATTTCTTAAAG 469  
QY 397 ACAAGATTTATCTTCTCAGCAAAATCACTTTTCAAGTCCATGCAATCTTCTATTTCAG 456  
DB 470 GGAAGACTTATCTTCTTAAAGTCTATTGATTTCAAGGCCCATGCAAAATCAATTAAGT 529  
QY 457 TAAAGATTTTGTGATCTTGAAGCAATCACTTTTCAAGTCCATGCAATCTTCTATTTCAG 513  
DB 530 TCCAGATCTTGAAGCACTTTATCAGCTTTCTACAAACATGAGATTTCAATGATGACAGA 589  
QY 514 GGCCTTGGATTGCTTTGATGATGTTCAAAATTTA-----GTGTTGAGAGAGAGAA 567  
DB 590 ACCACTGCTTATTTTGAGAGACGTTAATCTATCAATGATGAGGCGCTCGGCGGGA 649  
QY 568 CTATCAATGGCAATGACCAAGTATGTTGGCCAGTTCTTGCAAAATTAATATCACTGC 627  
DB 650 TTGTTGATGGCAAGGAAATCTGTGTGGCAAACTCAATGCAAAATCAAAATCTAAGC 709  
QY 628 CATGAGGGAATGACCAAGCGCTTAACTTCTGGAATGCAAAATTTGAAAGTA 687  
DB 710 CATGACAAAGCGCCCAAGCGCTCTACTCTCTACACCTTAAACATTTGATGAGGA 769  
QY 688 ATCTAAGATGAAAAATGACACAAATTCATCAATTAATGATGATGACTAATGTTG 747  
DB 770 ATCTGAGTGAAGAAATGACAGAGATTCAGATTTGATGAGAAATGCAACAGTGTG 829  
QY 748 TAGCTTCAATTTGATGATCAATGCTTACGAAAGGCCCAATCTGATGAGTCAATG 807  
DB 830 ATGTTAAGATGTTAAGATCACTGCTCGGCGATGTCACCAACGAGTGGTATTCATA 889  
QY 808 TATCAAAATCTCATATATTTCAATATCTGATACTATTTATGGAACAGTGTGATTTGA 867  
DB 890 TCGTTGCTACTAAAAACATTCGAATCTCAATTCAGACATTTGGGACAGGTGATGATGA 949  
QY 868 TTTCAATTTGTTCTGATCTCAAAATGTCAGGCCCAAAATTTACTGTGTGTCAGGTC 927  
DB 950 TATTCATTTGAGATGATGCAAAATGTTCAATCAATGATTTAACTTGCGGCCCGGTC 1009  
QY 928 ATGTTATAGTATTTGAAAGCTTTAGATCTGAAATTCAGAGCTTATGTCTAATGTTA 987  
DB 1010 ATGGCATGAGCATGTGAAAGCTTGGGGATGACAAATTCGCAAGCTTATGTATGGGAATTA 1069  
QY 988 CTGTAATGAAGCCAAATTTATGCGTCCGAAATGGAATGGAATCAAGACTTGGCAGG 1047  
DB 1070 ATGTGATGTGCTAGCGCTCTCTGAGACTGACATGAGTAAGATCAAGACTTACAGG 1129  
QY 1048 GAGGATCTGAGCAAGTCAACATCAATTTCTGATGTGAAATGCAAGCGTTAAGT 1107  
DB 1130 GAGGGTCAGGAACTGTTAAGACATTTAATTCGAAAATTCGTATGATGATCAAGA 1189  
QY 1108 ATCCCATTAATTAAGCAAACTATTGATGAGTTGAAACATGATACACAGTTT 1167  
DB 1190 ATCCGATCATATGACAGCACTACTGCGA---CAAGGCAAAATGCGAACAGCAGAAAT 1246  
QY 1168 CAGAGATTCAGTAAATAATGTGTATAGAAATCAAGGCAAGTGCACAAAGG 1227  
DB 1247 CTGGGTTCAAGTAAACATGCTGTATCTGAAACATTAAGGTCACAGCGCAACGATG 1306  
QY 1228 TGCGCATTAATTTGATTTGACGACAAACTTTCCATGTGAGGAATTAATGAGAAATA 1287  
DB 1307 TGCGGATTAATTTAATTTGAGTGTGAAATATCCATGCCCAAGGTATGTGCTGAGAAATG 1366



XX Rice gene, SEQ ID 3862.  
 XX Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN WO2003000898-A1.  
 PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Kakaigiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX  
 DR WPI; 2003-175290/17.  
 XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 PS Claim 6; SEQ ID NO 3862; 899bp; English.  
 XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC the expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 SQ Sequence 1269 BP; 373 A; 285 C; 319 G; 292 T; 0 U; 0 Other;  
 Query Match 22.9%; Score 375; DB 7; Length 1269;  
 Best Local Similarity 60.0%; Pred. No. 1.3e-56;  
 Matches 646; Conservative 0; Mismatches 425; Indels 6; Gaps 1;

Db 462 TGAAGGCGCAAGATTGTGGCGAATAATTCCTGCAAAACGAATGCAAACTTCATGCAC 521  
 QY 635 GGATGACCAAGGCGCTTAACTTCTGGAATTTGCAAAATTTGAAAGTGATATCTAA 694  
 Db 522 TGAAGCTCCAAAGCGCTTGACATTTCTCTGTTCAATCTGAAAGTGGAAGACTTGA 581  
 QY 695 GAGTAAATAATGCAACAAATTCATATCAATTTGAAGTCATGCATATGTTGAGCTTC 754  
 Db 582 GCTACTAAACAGCCAAACAAATCCACATGTGAGAGATTCAGTATGTTAGATCTC 641  
 QY 755 AAATTTGATATCATATCTTCAAGAAAGGCCAAATATGATGAGAGCCATGATCAAA 814  
 Db 642 TAGCTGACATACACACACAGGCACTAGCCCAACACACGCGCATTCATACAGAG 701  
 QY 815 TACTCAATATATTCAAATATCTGATATATTTATTTGAAACAGTGATGATTTTCAAT 874  
 Db 702 AAGTAAATAATGTAACATGACAGGCTGCATATATCAAGCCGGGATGACTCATGTGAT 761  
 QY 875 TGTTCGTGATCTCAAAATGTGACAGGCCCAAAATTTACTTGTGTCAGGTCAATGAT 934  
 Db 762 TGAAGATGAACTGAGAACTTACATGTCACAGACATGTGTGTGACCGGGACAGGAT 821  
 QY 935 AAGTATGGAAGCTTAGGATCTGAAATTCGAAGCTTATGTCTAATGTTACTGTAAA 994  
 Db 822 CAGATCGGAGGCTTAGGCGATCACTCTGAAAGCTCATTTCAAAATGTCACCGTGA 881  
 QY 995 TGAAGCCAAATTTATCGGTGCGCAAAATGAGATTAGATCAAGACTTGCGAGGAGATC 1054  
 Db 882 CACGTCAGGCTATATGSCACACACCAAGGAGCTCGCATCAAGACATGCGAGGCTG 941  
 QY 1055 TGAACAGCTAGCAACATCAAAATTTCTGAATGTGAAATTCGAAGCTTATGATTCAT 1114  
 Db 942 GGGCTACGCAAGAAACATCGTTCCAGACATGATCAAGAAATGTTGGAACCAAT 1001  
 QY 1115 AATATATACCAAAATCTATTTGATTCAGTTGAACCATGTATACACAGTTTTCAGCAGT 1174  
 Db 1002 CATATGACCAAGAACTATGATTTCTGTAACCTGTCAGAGAAACAGATCTGCGGT 1061  
 QY 1175 TCAAGTAAATAATGTGTGTATGAGAAATATCAAGGCACAAGTGCAACAAAGTGCGCAT 1234  
 Db 1062 GCAAGTTAGCAATGTGTCTTCAAGAACATCAAGGGGCAAGAGTCTTCCAAAGAGGCTAT 1121  
 QY 1235 AAATTTGATGACGACAAACTTTCCATGTGAAGATTTATAGGAATATATAA 1291  
 Db 1122 CAACTGAGCTGTAGCAAAATGTACTTGTCAAGAAATACCTTGAAAGAGCTGCA 1178

RESULT 13  
 ADA70540  
 ID ADA70540 standard; DNA; 1182 BP.  
 XX  
 AC ADA70540;  
 XX  
 AC 20-NOV-2003 (first entry)  
 XX  
 DE Rice gene, SEQ ID 3863.  
 XX  
 KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN WO2003000898-A1.  
 PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;



PI Katagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
 XX  
 DR WPI, 2003-175290/17.  
 XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 PS Claim 6; SEQ ID NO 3863; 899bp; English.  
 XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 SQ Sequence 1182 BP; 335 A; 275 C; 309 G; 263 T; 0 U; 0 Other;  
 Query Match 22.6%; Score 369; DB 7; Length 1182;  
 Best Local Similarity 60.1%; Pred. No. 1.4e-55;  
 Matches 634; Conservative 0; Mismatches 415; Indels 6; Gaps 1;  
 QY 263 TGGGATTAAAGTATTATGTTACTTACCTTGGAGCTTAAGGGTGAAGAAAACATATGA 322  
 DB 78 TGGCAGTAATGTGTTCAGCATACAGACTAOGGGGCTCATAGAGACGACGACATGATGA 137  
 QY 323 TAATATTGATTGAGACCATGGAATGAAGCATGTCATCTAGAACACCTGTTCAAT 382  
 DB 138 CACCAAGAATGGGAGATCATGGCTGACCTTGCTCTGCAAAACCTGACGTTT 197  
 QY 383 TGTGTTCTTAAACACAGATTTATCTTCACAGAAATACCTTTACAGTCCATGACG 442  
 DB 198 GCTCATCCCGAAGGAGAGAAATACCTGATCAAGCAACAACTGTGTGTCATGCA 257  
 QY 443 ATCTCTATTTCAGTAAAGATTTTGGATCCTTGAAGCATCTAGTAAATTTCAAGCT- 501  
 DB 258 ATCAAGCATCTCATGTATGTTGAAGGTTGTTGGTGCTTCCGACAGAGGTCACATG 317  
 QY 502 -----ACAAGATAGAAGCTTTGATGCTTTTGAATGTTCAAAATTTAGTTGG 556  
 DB 318 GAGCAAGAGACCATTAAGGACATGATCTGATCACTGATGTCATGCTTACCTGCAC 377  
 QY 557 AGGAGAGAGCACTATCATGGCATGGAAGACAGATAGTGGCCCAAGTTCTTGAATAAA 616  
 DB 378 TGGTGTGGGACCATATGATGGAATGCAAGATTGGTGGCAAAATTCATGCAAAACAA 437  
 QY 617 TAAATCACTGCGCATGCGAGGATGCAACACGCGCTTAACCTTCTGGAAATTCGCAAAATTT 676  
 DB 438 CTCAAAAGCTTCATGACAGAGAGCTCCAAAGGACACTGACATTCATCTCTGCAAGATCT 497  
 QY 677 GAAATGGAATATCTTAAAGATTAATAATGCAACAATTCATCAATCAATTTGAGTCATG 736  
 DB 498 GAAGGTAAGATCTGGAAGGTGATGAACAGCCAGCAAAATCCAGATTTCAGTGAAGATGG 557  
 QY 737 CACTAATGTTAGCTTCAAAATTTGATGATCATGTTTACAGAAAGAGCCCAATATCTGA 796  
 DB 558 CACCGATGATGATGATGCTCGCCCTGTCGATCAAGACCAAGAAACAGCCCCCAACATGA 617  
 QY 797 TGGAGTCATGATCAAAATCACTCAATATATTCATCAATCTGATCTATTTAGAACAGG 856  
 DB 618 TGGATCCATATCAACGACAGAGATGTTGAGTGAAGCTGATGATTAACACCGG 677  
 QY 857 TGAATATGTTAATTAATTTCTGATCTCAAAATGTCAGGCGCACAAATATTTACTTG 916  
 DB 678 GGATCACTGCAATGTAATGAGAGAACCCGAGAACCTGCAATGTCAGAAACATGCTGTG 737  
 QY 917 TGGTCCAGTCATGATTAAGTATGAAGCTTAGATCTGGAATTCAGAACTTAATGT 976

DB 738 TGACCCGGGACACGCGCATCAGCATGGTAGCTTGGGTGATCATTAATTTCTGAAGCTCATGT 797  
 QY 977 GTCTAATCTTCTGTAATGAAGCCAAATATATGTTGGTCCGAAATGAGTAGATCA 1036  
 DB 798 CAACAATGTCAACCTGCAATATGTCAGTTGTATAGCACACCAAGGAGCTGCGATCA 857  
 QY 1037 GACTTGGCAGGAGATCTGGAACAAGCTAGACAATCAAAATTTCTGAATGCGAAATGCA 1096  
 DB 858 GACATGGCAGGAGAGAAAGGTTTCAGCGAAGACATGCTGTTCCAGAAACATGGTCATGA 917  
 QY 1097 AGAGCTTAAGTATCCATTAATATATGACCAAACTAATTTGATGAGTGAACCATGAT 1156  
 DB 918 CAATGTCGGAACCCCATCATCATGACCAAACTAATGCACTTCTTACACCTTGCA 977  
 QY 1157 ACACAAGTTTTCAGCACTTCAAGTGAATAATGTGTTATGAGATATCAAGGCGACAG 1216  
 DB 978 GCAACGAAATCTGCAATGAGAGGTGAGACCAATGTCTGTTCAAGAAACATCAAGGCGACAG 1037  
 QY 1217 TGCAACAAGGTGGCCATTAATTTGATGAGCAACAACTTCCATGTGAAGAAATAT 1276  
 DB 1038 TGCATCAGAGAGAGGCCATCATGCTGCTGATGACGACGATGTACTTGCATGGCATTAAC 1097  
 QY 1277 AATGGAATATTAATTTAGTAGGGGAAAGTGGA 1311  
 DB 1098 CTGGAAGATGTCAATCTCACTGTCAAGGAGGA 1132  
 RESULT 14  
 AAV27950  
 ID AAV27950 standard; cDNA; 2207 BP.  
 XX  
 AC AAV27950;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 12-OCT-1998 (first entry)  
 XX  
 DE Polylacturonase genomic DNA.  
 XX  
 KM Transgenic plant; tissue-specific gene expression;  
 KM fruit-specific expression; polylacturonase; tomato; ss.  
 XX  
 OS Lycopersicon esculentum.  
 XX  
 PN US5750385-A.  
 XX  
 PD 12-MAY-1998.  
 XX  
 PF 07-JUN-1995; 95US-00484941.  
 XX  
 PR 17-JAN-1985; 85US-00692605.  
 PR 31-JUL-1986; 86US-00891528.  
 PR 26-MAY-1987; 87US-00054369.  
 PR 28-JUL-1987; 87US-00078538.  
 PR 25-JAN-1988; 88US-00147781.  
 PR 15-MAR-1988; 88US-00168190.  
 PR 29-APR-1988; 88US-00188361.  
 PR 07-NOV-1988; 88US-00257865.  
 PR 21-MAY-1990; 90US-00526123.  
 PR 09-JUL-1990; 90US-00550804.  
 PR 14-SEP-1990; 90US-00582241.  
 PR 08-AUG-1991; 91US-00742834.  
 PR 10-AUG-1993; 93US-00105852.  
 XX  
 PA (CALJ) CALGENE INC.  
 XX  
 PI Shewmaker CK, Facciotti D;  
 XX  
 DR WPI, 1998-296772/26.  
 XX  
 PT Producing transgenic plant of modified phenotype - is useful for  
 PT expression of heterologous genes which are light-inducible.  
 XX

PS Example 174; Fig 8A-C; 68pp; English.

XX This is the nucleotide sequence of a tomato polygalacturonase genomic DNA  
CC clone isolated from a genomic library by screening with polygalacturonase  
CC cDNA. The polygalacturonase gene promoter is active in at least the  
CC breaker through red fruit stage in tomato fruit, and can be used in a  
CC novel method of the invention for producing a plant with an altered  
CC phenotype. In this method, regulatory regions from genes expressed during  
CC a particular developmental stage or in a specific tissue are identified  
CC by cDNA screening. The regulatory regions are manipulated for use with  
CC foreign sequences for introduction into plant cells to provide  
CC transformed plants having a phenotypic property that can be modulated.  
CC The invention is exemplified with light, seed and fruit-specific  
CC promoters. Transformation without gall formation of cells which are not  
CC Agrobacterium hosts, especially cereals and legumes, is achieved.  
CC (Updated on 25-MAR-2003 to correct PF field.)

SQ Sequence 2207 BP; 870 A; 282 C; 270 G; 784 T; 0 U; 1 Other;

Query Match 20.3%; Score 331.8; DB 2; Length 2207;

Best Local Similarity 95.3%; Pred. No. 4.7e-49; Indels 0; Gaps 0;

Matches 342; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

QY 1 AATCTTTTCATGACAGGTTAAAAACCATACCATATACATATATCATGTTATCC 60
DB 1430 AATCTTTTCATGACAGGTTAAAAACCATACCATATACATATATCATGTTATCC 1489
QY 61 AAAGGAAATAGTATTCCTCTCTCATATATATTTTGGCTTCATCAATTCATGTTAGAA 120
DB 1490 AAAGGAAATAGTATTCCTCTCTCATATATATTTTGGCTTCATCAATTCATGTTAGAA 1549
QY 121 GCAATGTTATGATGACAAATTTATTCACAAAGTTATGATTAATTTCTTGAACAAGAT 180
DB 1550 GCAATGTTATGATGACAAATTTATTCACAAAGTTATGATTAATTTCTTGAACAAGAT 1609
QY 181 TTGCTCATGATTTTCAAGCTTATCTTCTTATTTTGGCAAAAATATGAAAGCAACATA 240
DB 1610 TTGCTCATGATTTTCAAGCTTATCTTCTTATTTTGGCAAAAATATGAAAGCAACATA 1669
QY 241 ATATTGACAGGTTGATTAATAAATGGGATTAAGGATTAATGATTAATGCTTTGGAGCTA 300
DB 1670 ATATTGACAGGTTGATTAATAAATGGGATTAAGGATTAATGATTAATGCTTTGGAGCTA 1729
QY 301 AGGGTGATGAAAAACATATGATTAATTTGATTTGAGACGATGAAGCATGT 359
DB 1730 AGGGTGATGAAAAACATATGATTAATTTGATTTGAGATTAATTTGATTTGAT 1788

```

# RESULT 15

AAV36973 standard; DNA; 2207 BP.

```

ID AAV36973
XX
AC AAV36973;
XX
DT 25-MAR-2003 (revised)
DT 26-OCT-1998 (first entry)
XX
DE Polygalacturonase genomic DNA.
XX
KW Transgenic plant; tissue-specific gene expression; promoter;
KW fruit-specific expression; tomato; polygalacturonase; ss.
OS Lycopersicon esculentum.
XX
PN US5753475-A.
XX
PD 19-MAY-1998.
XX
PF 10-AUG-1993; 93US-00105852.
XX
PR 17-JAN-1985; 85US-00692605.
PR 31-JUL-1986; 86US-00891529.
PR 26-MAY-1987; 87US-00054369.

```

PR 28-JUL-1987; 87US-00078538.  
PR 25-JAN-1989; 88US-00147781.  
PR 15-MAR-1988; 88US-00168190.  
PR 29-APR-1988; 88US-00188361.  
PR 02-NOV-1988; 88US-00267685.  
PR 21-NOV-1990; 90US-00526123.  
PR 09-JUL-1990; 90US-00550804.  
PR 14-SEP-1990; 90US-00582241.  
PR 08-AUG-1991; 91US-00742834.

XX (CALJ ) CALGENE INC.

XX Houck CM;

XX WPI; 1998-311403/27.

PT Transformation of plants - with regulatory sequence containing constructs  
PT for tissue specific expression of genes.

PS Example 17; Fig 8A-C; 67pp; English.

XX This is the nucleotide sequence of a tomato polygalacturonase genomic DNA  
CC clone isolated from a genomic library by screening with polygalacturonase  
CC cDNA. The polygalacturonase gene promoter is active in at least the  
CC breaker through red fruit stage in tomato fruit, and can be used in a  
CC novel method of the invention for producing a plant with a regulatable  
CC phenotypic property. In this method, regulatory regions from plant genes (and T-DNA  
CC and Ti or Ri plasmids) are manipulated for use with foreign sequences for  
CC introduction into plant cells to provide transformed plants having a  
CC phenotypic property that can be modulated. The invention is exemplified  
CC with light, seed and fruit-specific promoters. Also claimed are methods  
CC for altering the phenotype of fruit tissue as distinct from other plant  
CC tissue, modifying the genotype of a plant to impart a desired  
CC characteristic to fruit, modifying transcription in fruit tissue, and  
CC expressing a heterologous DNA sequence of interest specifically in fruit  
CC tissue. The method provides transformation without gall formation in  
CC plants which have historically not been Agrobacterium hosts. (Updated on  
CC 25-MAR-2003 to correct PF field.)

SQ Sequence 2207 BP; 870 A; 282 C; 270 G; 784 T; 0 U; 1 Other;

Query Match 20.3%; Score 331.8; DB 2; Length 2207;

Best Local Similarity 95.3%; Pred. No. 4.7e-49; Indels 0; Gaps 0;

Matches 342; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

QY 1 AATCTTTTCATGACAGGTTAAAAACCATACCATATACATATATCATGTTATCC 60
DB 1430 AATCTTTTCATGACAGGTTAAAAACCATACCATATACATATATCATGTTATCC 1489
QY 61 AAAGGAAATAGTATTCCTCTCTCATATATATTTTGGCTTCATCAATTCATGTTAGAA 120
DB 1490 AAAGGAAATAGTATTCCTCTCTCATATATATTTTGGCTTCATCAATTCATGTTAGAA 1549
QY 121 GCAATGTTATGATGACAAATTTATTCACAAAGTTATGATTAATTTCTTGAACAAGAT 180
DB 1550 GCAATGTTATGATGACAAATTTATTCACAAAGTTATGATTAATTTCTTGAACAAGAT 1609
QY 181 TTGCTCATGATTTTCAAGCTTATCTTCTTATTTTGGCAAAAATATGAAAGCAACATA 240
DB 1610 TTGCTCATGATTTTCAAGCTTATCTTCTTATTTTGGCAAAAATATGAAAGCAACATA 1669
QY 241 ATATTGACAGGTTGATTAATAAATGGGATTAAGGATTAATGATTAATGCTTTGGAGCTA 300
DB 1670 ATATTGACAGGTTGATTAATAAATGGGATTAAGGATTAATGATTAATGCTTTGGAGCTA 1729
QY 301 AGGGTGATGAAAAACATATGATTAATTTGATTTGAGACGATGAAGCATGT 359
DB 1730 AGGGTGATGAAAAACATATGATTAATTTGATTTGAGATTAATTTGATTTGAT 1788

```

Search completed: July 26, 2004, 23:42:35  
Job time : 674 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 23:31:25 ; Search time 127 Seconds

(without alignments)  
7148.816 Million cell updates/sec

Title: US-10-018-604-1

Sequence: 1 aactcttttcaatagacag.....atatgaaaaaaaaaaaaa 1636

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	1636	6 5447867-2	Patent No. 5447867
2	420.4	25.7	1657	3 US-08-941-532-5	Sequence 5, Appl1
3	420.2	25.7	1631	4 US-09-051-239A-1	Sequence 1, Appl1
4	228	13.9	1410	3 US-08-467-023-139	Sequence 139, Appl
5	228	13.9	1479	3 US-08-467-023-141	Sequence 141, Appl
6	228	13.9	1726	3 US-08-467-023-133	Sequence 133, Appl
7	225.4	13.8	1395	3 US-08-467-023-140	Sequence 140, Appl
8	212	13.0	219	1 US-08-463-213-1	Sequence 1, Appl1
9	212	13.0	219	6 5453566-1	Patent No. 5453566
10	73	4.5	164	3 US-08-941-532-7	Sequence 7, Appl1
11	58.4	3.6	7218	1 US-08-232-463-14	Sequence 14, Appl1
12	53	3.2	640681	4 US-09-790-988-1	Sequence 1, Appl1
13	52.4	3.2	1356	4 US-09-107-532A-2998	Sequence 2998, Ap
14	52	3.2	640681	4 US-09-790-988-1	Sequence 1, Appl1
15	51.8	3.2	6866	4 US-10-204-708-19	Sequence 19, Appl1
16	51.6	3.2	53832	4 US-09-801-861-3	Sequence 3, Appl1
17	51.2	3.1	832	4 US-09-621-976-2813	Sequence 2813, Ap
18	50.4	3.1	11049	4 US-10-204-708-22	Sequence 22, Appl1
19	50	3.1	19124	2 US-08-487-826B-13	Sequence 13, Appl1
20	49.6	3.0	665	4 US-08-883-795A-36	Sequence 36, Appl1
21	49.6	3.0	53332	4 US-09-801-861-3	Sequence 3, Appl1
22	48.8	3.0	6152	3 US-08-973-462-1	Sequence 1, Appl1
23	48.2	2.9	5340	4 US-09-627-122-21	Sequence 21, Appl1
24	47.6	2.9	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
25	47.6	2.9	1636	6 5447867-2	Patent No. 5447867
26	47.6	2.9	20674	4 US-09-641-638-651	Sequence 651, Appl
27	47.2	2.9	2663	1 US-08-136-743B-3	Sequence 3, Appl1

28	47.2	2.9	3412	3 US-08-903-325-3	Sequence 3, Appl1
29	47.2	2.9	8920	2 US-08-446-855A-1	Sequence 1, Appl1
30	47.2	2.9	8920	3 US-09-150-741-1	Sequence 1, Appl1
31	47.2	2.9	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
32	47	2.9	2217	3 US-09-244-314-1	Sequence 1, Appl1
33	47	2.9	2217	3 US-09-498-959-1	Sequence 7, Appl1
34	46.8	2.9	155	4 US-09-051-239A-7	Sequence 27, Appl1
35	46.6	2.8	4818	3 US-08-817-926-27	Sequence 520, Appl
36	46.2	2.8	3982	4 US-08-956-171E-520	Sequence 10, Appl1
37	46.2	2.8	6070	4 US-10-204-708-10	Sequence 28, Appl
38	46	2.8	11131	4 US-10-204-708-10	Sequence 24, Appl1
39	45.6	2.8	688	3 US-08-998-416-972	Sequence 1, Appl1
40	45.4	2.8	2110	3 US-09-419-459-1	Sequence 1, Appl1
41	45.4	2.8	11049	4 US-10-204-708-24	Sequence 1, Appl1
42	45.2	2.8	2861	1 US-08-299-953-1	Sequence 1, Appl1
43	45.2	2.8	2861	1 US-08-459-415-1	Sequence 1, Appl1
44	45.2	2.8	2861	4 US-09-066-687-1	Sequence 1, Appl1
45	45.2	2.8	2861	5 PCT-US95-11231-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1									
5447867-2									
Patent No. 5447867									
TITLE OF INVENTION: BRIDGES, IAN;SCHUCH, WOLFGANG;GRISON, DONALD									
REPERASE GENE SEGMENTS									
NUMBER OF SEQUENCES: 4									
CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US /08/24, 866									
FILING DATE: 26-FEB-1993									
PRIOR APPLICATION DATA:									
APPLICATION NUMBER: 720,629									
FILING DATE: 25-JUN-1991									
APPLICATION NUMBER: 419,779									
FILING DATE: 29-SEP-1989									
APPLICATION NUMBER: 119,614									
FILING DATE: 12-NOV-1987									
SEQ ID NO:2:									
LENGTH: 1636									
5447867-2									
Query Match									
Best Local Similarity 100.0%; Pred No. 0;									
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	AACTCTTTTCAATAGACAGTTTAAACCACATATACATATACATATATCATGTTATCC	60						
DB	1	AACTCTTTTCAATAGACAGTTTAAACCACATATACATATACATATATCATGTTATCC	60						
QY	61	AAAGGAATGATTTCTCTCTTCATTTATTTTGGCTTCATCATTTCACTTGAGAA	120						
DB	61	AAAGGAATGATTTCTCTCTTCATTTATTTTGGCTTCATCATTTCACTTGAGAA	120						
QY	121	GGAATGTTATGATGACAAATTTTCAACAAAGTTTATGATATATTTCTTGAACAGAT	180						
DB	121	GGAATGTTATGATGACAAATTTTCAACAAAGTTTATGATATATTTCTTGAACAGAT	180						
QY	181	TTGCTCATGATTTTCAAGCTTATCTTTCTTATTTGACAAATTTTGAACACATTA	240						
DB	181	TTGCTCATGATTTTCAAGCTTATCTTTCTTATTTGACAAATTTTGAACACATTA	240						
QY	241	ATATTGACAAAGTTTAAATGGAATTAAGTATTAATGTTACTTACCTTGGAGCTA	300						
DB	241	ATATTGACAAAGTTTAAATGGAATTAAGTATTAATGTTACTTACCTTGGAGCTA	300						
QY	301	AGGGTATGGAACAAATATGATTAATTTGATTTGACAAAGCAAGCAATGATGTT	360						
DB	301	AGGGTATGGAACAAATATGATTAATTTGATTTGACAAAGCAAGCAATGATGTT	360						
QY	361	CATCTAGAACACCTGTTCAATTTTGGTTCTCTTAAACAAAGATTAATTTCTTGAACAA	420						

DB 361 CATCTAGAACCCCTGTCATTTGTGGTCCCTAAACCAAGAAATATCTTCTCAAGCAAA 420  
QY 421 TCACCTTTGAGGTCATGAGATCTTCTATTTCAATTAAGATTTTGGATTCCTTAGAAG 480  
DB 421 TCACCTTTGAGGTCATGAGATCTTCTATTTCAATTAAGATTTTGGATTCCTTAGAAG 480  
QY 481 CATCTAGTAAATTTGACACTACCAAGATGAAAGGCTTTGGATTTGCTTTGATAGTTC 540  
DB 481 CATCTAGTAAATTTGACACTACCAAGATGAAAGGCTTTGGATTTGCTTTGATAGTTC 540  
QY 541 AAAATTTAGTTGGGAGAGAGAGAACTATGCAATGCAATGCAAGTATGGTGGCCAA 600  
DB 541 AAAATTTAGTTGGGAGAGAGAGAACTATGCAATGCAATGCAAGTATGGTGGCCAA 600  
QY 601 GTTCTTCAAAATTAATTAATCACTGCGCATGCAAGGATGCAACCAAGGCTTTAACTTCT 660  
DB 601 GTTCTTCAAAATTAATTAATCACTGCGCATGCAAGGATGCAACCAAGGCTTTAACTTCT 660  
QY 661 GGAATTCGCAAAATTTGAAAGTGAATATCTAAAGATGAAATGCAACCAATTCATA 720  
DB 661 GGAATTCGCAAAATTTGAAAGTGAATATCTAAAGATGAAATGCAACCAATTCATA 720  
QY 721 TCAATTTGAGTCACTGACCTAAATGTTGATGCTCAAAATTTGATGATGAATGCTTCAAGAA 780  
DB 721 TCAATTTGAGTCACTGACCTAAATGTTGATGCTCAAAATTTGATGATGAATGCTTCAAGAA 780  
QY 781 AGAGCCCAAAATCTGATGAGTCCATGATCAATCAATCAATCAATCAATCAATCAATCA 840  
DB 781 AGAGCCCAAAATCTGATGAGTCCATGATCAATCAATCAATCAATCAATCAATCAATCA 840  
QY 841 CTATTTATGGAACAGGTGATGATTTGATTTCAATTTGCTGAGTCTCAAAATGTGACAG 900  
DB 841 CTATTTATGGAACAGGTGATGATTTGATTTCAATTTGCTGAGTCTCAAAATGTGACAG 900  
QY 901 CCACAAATATTTACTTGGTCCAGGTCATGATTAAGATTAAGATTAAGATTAAGATTTGAA 960  
DB 901 CCACAAATATTTACTTGGTCCAGGTCATGATTAAGATTAAGATTAAGATTAAGATTTGAA 960  
QY 961 ATTCAAGAGTTATGCTCTATATGTTACTGTAATTAAGAACCAAAATTTATGCGTCCGAAA 1020  
DB 961 ATTCAAGAGTTATGCTCTATATGTTACTGTAATTAAGAACCAAAATTTATGCGTCCGAAA 1020  
QY 1021 ATGAGTTAGATCAAGCTTGGCAGGAGGATCTGGAACAAGCTTAAGCAAAATTTTC 1080  
DB 1021 ATGAGTTAGATCAAGCTTGGCAGGAGGATCTGGAACAAGCTTAAGCAAAATTTTC 1080  
QY 1081 TGAATGTGGAATGCAAGCTTGAATCCATTAATTAAGAACCAAAATTTATGCGTCCGAAA 1140  
DB 1081 TGAATGTGGAATGCAAGCTTGAATCCATTAATTAAGAACCAAAATTTATGCGTCCGAAA 1140  
QY 1141 GAGTTGAACCATGTATACACAGTTTTCAGCAGTTCAAGTGAAGAAATGTGTGTATGAGA 1200  
DB 1141 GAGTTGAACCATGTATACACAGTTTTCAGCAGTTCAAGTGAAGAAATGTGTGTATGAGA 1200  
QY 1201 ATATCAAGGAGCAAGCTGCAACAAAGGTGGCCATTAATTTGATGCAAGCAAAATTTTC 1260  
DB 1201 ATATCAAGGAGCAAGCTGCAACAAAGGTGGCCATTAATTTGATGCAAGCAAAATTTTC 1260  
QY 1261 CATGTGAAGGAATTTAATGAGAAATATTAATTTGATGAGGAAAGTGGAAACCATCAG 1320  
DB 1261 CATGTGAAGGAATTTAATGAGAAATATTAATTTGATGAGGAAAGTGGAAACCATCAG 1320  
QY 1321 AGGCTACGTCGCAAAATGTCATTTTAACATGTCGAAATGTCACCAACATGCACTT 1380  
DB 1321 AGGCTACGTCGCAAAATGTCATTTTAACATGTCGAAATGTCACCAACATGCACTT 1380  
QY 1381 CACTAGAAATTTCAAGAGTGAAGCTTTTGTATATTAATTAATTAATTAATTAATTAAT 1440  
DB 1381 CACTAGAAATTTCAAGAGTGAAGCTTTTGTATATTAATTAATTAATTAATTAATTAAT 1440  
QY 1441 TCAATATTAAGAGATATATATATATCAATTAACAAATCTATATCTATGATTTGAAATA 1500  
DB 1441 TCAATATTAAGAGATATATATATATCAATTAACAAATCTATATCTATGATTTGAAATA 1500

DB 1441 TCAATATTAAGAGATATATATATATCAATTAACAAATCTATATCTATGATTTGAAATA 1500  
QY 1501 TTAATTAATTAATGATGAGATTTGAGTTTAATTAAGATTAATTAATTAATTTCTA 1560  
DB 1501 TTAATTAATTAATGATGAGATTTGAGTTTAATTAAGATTAATTAATTAATTTCTA 1560  
QY 1561 GTCAAAAGTTTGAAGATTTGATTTTAATGATCAAAATTAATTAATTAATTAATTAAT 1620  
DB 1561 GTCAAAAGTTTGAAGATTTGATTTTAATGATCAAAATTAATTAATTAATTAATTAAT 1620  
QY 1621 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1636  
DB 1621 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1636

RESULT 2  
US-08-941-532-5  
Sequence 5, Application US/08941532  
Patent No. 6096946  
GENERAL INFORMATION:  
APPLICANT: ROBERTS, Jeremy Alan  
APPLICANT: COUPE, Simon Allan  
APPLICANT: JENKINS, Elizabeth Sarah  
TITLE OF INVENTION: CONTROL OF POD DEHISCENCE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,532  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/00757  
FILING DATE: 29-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9506684.1  
FILING DATE: 31-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0623.0580001/RWE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1657 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 145..1446  
US-08-941-532-5

Query Match 25.7%; Score 420.4; DB 3; Length 1657;  
Best Local Similarity 62.5%; Pred. No. 2.8e-87;  
Matches 737; Conservative 0; Mismatches 416; Indels 27; Gaps 4;

QY 277 TTAATGTAATTAAGCTTTGAGGCTTAAGGATGAGAAACATATGATTAATTAATTAAT 336  
DB 350 TTAAGTTTGAACCTTGGAGGCAAAAGGTGATGGAACCAACCGATGATTAATTAATTAAT 409

QY 337 AGCAAGATGAGTGAAGATGCTTACCTAGAACCTGTTCAATTTGCTGCTTAAA 396  
 Db 410 AGAAGATGGAAGAGGATGCTTCAACAAATGAGTGAAGTACTTTCTTGATTCCTAAAG 469  
 QY 397 ACAGAAATTAATCTCTCAAGCAATCACTTTGAGGCTCAAGAGATCTTCTATTTCAG 456  
 Db 470 GGAAGACTTAATCTCTTAAGCTTATTAAGTTTCAAGAGGCTCAAGATTCATTAAGTACT 529  
 QY 457 TAAAGATTTTGGATCTTGAAGCAATCTAGTAAATTTTGAAGTAC---AAAGATGAA 513  
 Db 530 TCCAGATCTAGGCACTTTATCAAGCTTCTAATAAATCATGGATTAAGTAAATACAGA 589  
 QY 514 GGGTTTGGATTCCTTTGATAGTGTCAAAATTTA-----GTTGTTGAGGAGAGAGAA 567  
 Db 590 ACCACTGGCTTATTTTGGAGAGCTTAAATCATCATCATGATGAGGCTCGGCGGGGGA 649  
 QY 568 CTATCATGAGCAATGAGCAAGTATGAGGCAAGTTTTCACAAATTAATTAATCACTGC 627  
 Db 650 TTGTTGATGCAACGAGAAAATCTGTTGCAAAATCTATCAAAATTCGACAAATCTAAGC 709  
 QY 628 CATGCAAGGATGCAACAGGCTTAACTTTGGAATTCGAAAAATTTGAAAGTGAATA 687  
 Db 710 CATGCAAAAGGCGCAAGGCTTAACTTCTTAACAACCTAAACATTTGAATGTGAGA 769  
 QY 688 ATCTAAAGATTAATTAATGCAACAAATTTATTAATTTGATGATGCAATGATGTTG 747  
 Db 770 ATCTGAGAGTGAAGAAATGACAGCAATTCAGATTTGATGGAATGCAAGATGTTG 829  
 QY 748 TAGCTTCAATTTGATGATCAATGCTTCAAGCAAGGCGCAATTAATGATGAGTGCAG 807  
 Db 830 ATCTTAAGATGTTAAGATCACTGCTCTGGCGATGCTCCACAGGAGATTCATA 889  
 QY 808 TATCAAAATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 867  
 Db 890 TCGTTCTACTAAATCAATTCGATCTCAATTCACAAATTTGGAACAGTGAATGCA 949  
 QY 868 TTTTCAATTTTCTGATCTCAAAATGTCAGGCGCAAAATTTACTTGTGTCAGGTC 927  
 Db 950 TATCCATTAAGATGATGCAAAATGTTCAATCATGATTTAACTTGGCGCCGCGTC 1009  
 QY 928 ATGTAATAGTATTTGGAAGCTTGAAGTCTGGAATTCAGAAAGTTATGCTCAATGTTA 987  
 Db 1010 ATGGATCTGCAATTTGGAAGCTTGGGAGATGACATTTCAAGTTATGATGGGAATTA 1069  
 QY 988 CTGTAATGAAGCCAAATTAATTCGTGCGGAAATGAGTTGATCAAGACTTGGCAG 1047  
 Db 1070 ATGTGATGCTGCTACGCTCTCTGAGACTGACATGAGATTAAGATCAAGACTTACAG 1129  
 QY 1048 GAGGATCTGGAACAGTGAAGCAATCAATTTTCTGATGATGGAATGCAAGCGTTAACT 1107  
 Db 1130 GAGGATCTGGAACAGTGAAGCAATCAATTTTCAAAATTCATGATGATGATGCAAGA 1189  
 QY 1108 ATCCATTAATTTAGAACCAAACTATTTGATGAGTTGTAACCATGATTAACAAGCTTT 1167  
 Db 1190 ATCCATTAATTTAGAACCAAACTATTTGAGCTGCA---CAAGCAAAATGCGAACAAGAT 1246  
 QY 1168 CAGCAATTAAGTGAAGTGTGATGAGATTAACAAGGCAAGTGAAGCAAAAG 1227  
 Db 1247 CTGCGATTAAGTGAAGTGTGATGAGATTAACAAGGCAAGTGAAGCAAAAG 1306  
 QY 1228 TGGGATTAATTTGATGAGCAAACTTTCTGATGAGAGAAATTAATGAGAAATA 1287  
 Db 1307 TGGGATTAATTTGATGAGCAAACTTTCTGATGAGAGAAATTAATGAGAAATA 1366  
 QY 1288 TAAATTTAGTGAAGGAGTGAAGAACCATAGAGGCTAGCGGCAAAATGCGCAATTTTA 1347  
 Db 1367 TGAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1411  
 QY 1348 ACAATGCTGAACATGTTACACCACTGCACTGCAATTTCAAGAGATGAGAGCTC 1407  
 Db 1412 AGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1471

QY 1408 TTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1447  
 Db 1472 ATAAATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1511  
 RESULT 3  
 US-09-051-239A-1  
 / Sequence 1, Application US/09051239A  
 / Patent No. 6420628  
 / GENERAL INFORMATION:  
 / APPLICANT: ULVSKOV, Peter  
 / APPLICANT: CHILD, Robin  
 / APPLICANT: VAN ONCKELIN, Henri  
 / APPLICANT: PRINSEN, Els  
 / APPLICANT: BORKHARDT, Bernard  
 / APPLICANT: SANDER, Lilli  
 / APPLICANT: PETERSEN, Morten  
 / APPLICANT: BONGARD POUSEN, Gert  
 / APPLICANT: BOTTERMAN, Johan  
 / TITLE OF INVENTION: Seed Shattering  
 / FILE REFERENCE: 2121-0138P  
 / CURRENT APPLICATION NUMBER: US/09/051,239A  
 / PRIOR FILING DATE: 1998-09-28  
 / PRIOR APPLICATION NUMBER: PCT/EP96/04313  
 / PRIOR FILING DATE: 1996-10-04  
 / PRIOR APPLICATION NUMBER: EP 95 402241.4  
 / PRIOR FILING DATE: 1995-10-06  
 / PRIOR APPLICATION NUMBER: EP 95 203328.0  
 / PRIOR FILING DATE: 1995-12-08  
 / NUMBER OF SEQ ID NOS: 14  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 1  
 / LENGTH: 1631  
 / TYPE: DNA  
 / ORGANISM: Brassica napus  
 / FEATURE:  
 / OTHER INFORMATION: Location 95-163 = region encoding the presumed  
 / OTHER INFORMATION: endo-PG signal peptide.  
 / OTHER INFORMATION: Location 884-900 = region of the endo-PG cDNA  
 / OTHER INFORMATION: corresponding to oligonucleotide PG3  
 / OTHER INFORMATION: Location 1059-1073 = region of the endo-PG cDNA  
 / OTHER INFORMATION: complementary to oligonucleotide PG2  
 / OTHER INFORMATION: Location 1229-1245 = region of the endo-PG cDNA  
 / OTHER INFORMATION: complementary to oligonucleotide PG5  
 / NAME/KEY: CDS  
 / LOCATION: (95)..(1393)  
 / OTHER INFORMATION: Location 821-837 = region of endo-PG cDNA  
 / OTHER INFORMATION: corresponding to oligonucleotide PG1.  
 / OTHER INFORMATION: Strain cv. Topaz.  
 / NAME/KEY: unsure  
 / LOCATION: (1439)  
 / OTHER INFORMATION: n = a, c, g, t, any, other, unknown, or other  
 US-09-051-239A-1  
 Query Match 25.7%; Score 420.2; DB 4; Length 1631;  
 Best Local Similarity 63.6%; Pred. No. 3,1e-87;  
 Matches 717; Conservative 0; Mismatches 383; Indels 27; Gaps 4;  
 QY 277 TTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 336  
 Db 300 TTAATGTTTGAAGTCTGAGGCAAGGATGAGAAATCGAATGATGATGATGATGATGATG 359  
 QY 337 AGCAAGCATGGAATGAAGCATGTTGATGAGCAAGCTGTTCAATTTGAGTTGCTTAAAG 419  
 Db 360 AGAAGCATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 419  
 QY 397 ACAGAAATTAATCTCTCAAGCAATCACTTTTGAAGTCAAGTCAATTTGATTTGAG 456  
 Db 420 GAAAGACTTAATCTCTTAAGCTTATTAAGTTTCAAGAGGCTCAAGTCAATTTGATGAGT 479  
 QY 457 TAAAGATTTTGGATCTTGAAGCAATCTAGTAAATTTTGAAGTAC---AAAGATGAA 513  
 Db 480 TCCAGATCTAGGCACTTTATCAAGCTTCTAACAAGATCGGATTAAGTAAATGAGCAAGA 539

QY 514 GGCCTTGAGTCTTTGATGATGTTCAAAATTTA-----GTTGTGAGAGAGAGAA 567  
DB 540 ACCACTGGCTTATTTGGAGAGCTTAATATATCATCATGATGAGCGGCTCGCGGAGAA 599  
QY 568 CTATCAATGGCAATGACAGATATGAGGCAAGTTCTTGCAAAATTAATTAATCTACTGC 627  
DB 600 TTGTTGATGGCAAGAGAAATATCTGTGGGCAAAATCATGCAAAATCTTAAGC 659  
QY 628 CATGCAAGGATGACCAAGCGCTTAACTTTGGAATTGCAAAATTTGAAAGTGATA 687  
DB 660 CATGCAAAAAGCGCCCAAGCGCTTCTCTCAACAATAAGAAATTTGATGTGAAGA 719  
QY 688 ATCTAAAGATGAAAATGACCAACAATTCATATCAATTTGAGTCATGACCTAATGTTG 747  
DB 720 ATCTGAAGATGAGAAATGACAGAGATTCAGATTTGATGAAATTTGCAACAATGTTG 779  
QY 748 TAGCTTCAAAATTTGATGATCATGCTTCAAGCAAAAGCGCCCAATATCTGATGAGTCCATG 807  
DB 780 GCGTTAAGATGTTAAGATCTCTCTCTGCGCATATGTCACACAGGATGTTATTCATA 839  
QY 808 TATCAAAATCTCAATATATTCATATCTGATTAATTTTGGAAACAGGTATGATTTGTA 867  
DB 840 TCGTTGCTACTAAATAATTCGAATTCGAATTCAGACATTTGAGACAGGTATGATTTGTA 899  
QY 868 TTTCAATTTGTTCTGAGATCTCAAAATGTCAGGCAAAATATTAATTTGTCAGGTC 927  
DB 900 TATCATTTGAGAGATGATGCAAAATGTTCAATCAATGATTTAACTTTGCGGCCCCGCTC 959  
QY 928 ATGTATATGATTTGGAAGCTTAGATCTGAAATTCGAAGCTTATGTCTAATGTTA 987  
DB 960 ATGGATTCAGATTTGGAAGCTTTGGGCGGATGCAATTCGAAGCTTATGTATCGGAATTTG 1019  
QY 988 CTGTAATGAAAGCCAAATTTATCGTCCGCAAAATGAGTTAGATCAAGCTTTGGCAGG 1047  
DB 1020 ATGTGATGATGCTAAGCTCTCTGAGATGCAATGAGTAAAGATCAAGCTTTACAGG 1079  
QY 1048 GAGGATCTGCAACAGCTGCAACATTCGAATTTGGAATGGAATGCAAGCTTTAGT 1107  
DB 1080 GAGGATCTGCAACAGCTGCAACATTCGAATTTGGAATGGAATGCAAGCTTTAGT 1139  
QY 1108 ATCCATATATATAGACCAAACTATTTGATGATGGAATGGAATGATACACAGTTT 1167  
DB 1140 ATCCATATATATAGACCAAACTATTTGATGATGGAATGGAATGATACACAGTTT 1196  
QY 1168 CAGCAATGCAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1227  
DB 1197 CTGCGGTTCAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1256  
QY 1228 TGGCATTAATTTGATTTGAGCAGCAAACTTTCCATGGAAGATTTAATGAGATTA 1287  
DB 1257 TGGCATTAATTTGATTTGAGCAGTGTGAAATATTCATGCGCAAGTATTTGCTTGAGATG 1316  
QY 1288 TAAATTTAGTAGGGAATGGAATAATTCAGAGGCTACGTGCAAAATTTGCTAATTTTA 1347  
DB 1317 TGAACATCAAAAGAGAA-----AAGCTTCTGCAAAATGCTCAATGTTA 1361  
QY 1348 ACAATGCTGAACATGTTACACACACTGATCTTCACTAGAAATTTCA 1394  
DB 1362 AGGATTAAGGCACTGTTCTCTTAATGCTTAATTTACTAAGCTGA 1408

RESULT 4  
US-08-467-023-139

/ Sequence 139, Application US/08467023

/ Patent No. 6090386

/ GENERAL INFORMATION:

/ APPLICANT: Griffith, Irwin J.;

/ APPLICANT: Pollock, Joanne;

/ APPLICANT: Bond, Julian F.;

/ APPLICANT: Garman, Richard D.;

/ APPLICANT: Kuo, Mei-Chang;

/ APPLICANT: Yeung, Siu-mei H.;

APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Maltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 139:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-467-023-139

Query Match 13.9%; Score 228; DB 3; Length 1410;  
Best Local Similarity 52.3%; Pred. No. 3.4e-43;  
Matches 557; Conservative 0; Mismatches 495; Indels 12; Gaps 2;  
QY 240 AATATTGACAGGTGATATAAAATGAGTTAAAGTATTAATGATCTTAGCTTTGAGCT 299  
DB 4 AAAGTTGAGCAATCTCGTCATGATCTATCAACATCTTCAATGTGAAAGATATGCGCA 63  
QY 300 AAGGATGAGAAACATATGATTAATTTGATTTGAGCAAGCATGCAATGAACATGT 359  
DB 64 GTAGGCGATGGAACATGATGATGACATGAGGATTTTCAACGCAATGCGAAGCTGATGC 123  
QY 360 TCATCTAGAACACCGTTCAATTTGAGTCTTAATAAACAGAAATTAATCTTCAAGCA 419  
DB 124 AAAAACCATGCAATGATGCTTGTG--CGAGCAGCAAGAAATTTGTTTAACAT 180  
QY 420 ATCACTTTTCAAGTCAATGACATCTTCTATTTTCAAGTAAGATTTTGGATCTTAGAA 479  
DB 181 CTGTTCTTCAATGAGGCAATGCACTCACTTACTTTAAGATGATGGAATTAATAGCT 240  
QY 480 GCATCTAGTAATTTCAAGCTTCAAAAGTATGAAGCTTTGATTTGCTTTGATAGTGT 539  
DB 241 GCGTACCAAAATCCAGCGAGCTGGAAGATATATGAATATGTTGAGCTTCTAAACTT 300  
QY 540 CAAATTTTATGTTTGGAGGAGAGAACTATCATGCGAATGCAACATATGATGAGCA 599  
DB 301 ACAGGTTTACTCTAATGAGTAAGGTTAATGATGAGCAAGAAACAAATGATGAGCT 360  
QY 600 AGTCTTGCATAATTAATTAATCACTGCCA-----TGAGGAGATCACCACAGGCC 650  
DB 361 GGCATATGTAATGGTCAATGAGCAAGAAATTTGCAAGATGATGATGAGCAACAGGCC 420

1	SOFTWARE: Patentin Release #1.0, Version #1.25				
2	CURRENT APPLICATION DATA:				
3	APPLICATION NUMBER: US/08/467,023				
4	FILING DATE: June 6, 1995				
5	CLASSIFICATION: 424				
6	PRIOR APPLICATION DATA:				
7	APPLICATION NUMBER: 08/350,225				
8	FILING DATE: December 6, 1994				
9	ATTORNEY/AGENT INFORMATION:				
10	NAME: Jane E. Remillard				
11	REGISTRATION NUMBER: 38,872				
12	REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)				
13	TELECOMMUNICATION INFORMATION:				
14	TELEPHONE: (617) 227-7400				
15	TELEFAX: (617) 227-5941				
16	INFORMATION FOR SEQ. ID NO: 141:				
17	SEQUENCE CHARACTERISTICS:				
18	LENGTH: 1479 base pairs				
19	TYPE: nucleic acid				
20	STRANDEDNESS: single				
21	TOPOLOGY: linear				
22	MOLECULE TYPE: CDNA				
23	US-08-467-023-141				
24					
25	Query Match	13.9%	Score 228;	DB 3;	Length 1479;
26	Best Local Similarity	52.3%	Pred. No. 3.4e-43;		
27	Matches 557; Conservative	0;	Mismatches 495;	Indels 12;	Gaps 2
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					
90					
91					
92					
93					
94					

```

Db      730 AATATTGTGATTGAGATTCGATTTGGGCTCCAGCCATGATTAAGTATAGGAAGCTT 789
QY      951 GGATCTGGAATTCAGAGCTTATGTGCTAATGTTACTGTAATGAGCCAAAATATTC 1010
Db      790 GGGAGGGGAAAACCTAGACAGAGAGTTTCACTGACGTAATGAGGCTAAATTCATA 849
QY      1011 GGTGCCAAAATGAGTTAGATCAAGATTCGGAGGAGAGATCTGACAAAGCTAGCAAC 1070
Db      850 GACACACAAAATGATTAAGATCAAAACATGGAGGATGTTAGGCAATGGCAAGCCAT 909
QY      1071 ATCAAAATTCGATGATGGAATGCAAGAGCTTAAGTATCCCATTAATTAAGACCAAAAC 1130
Db      910 ATATATTATAGAAATGTGAAATGATAAATTCGAGAAACCCATATTAATTAATCAATTC 969
QY      1131 TATTGTATGAGTTGAACCATGTATACACAGTTTTCAGAGTTCAAGGAAAAATGTG 1190
Db      970 TACTGCATTCAGCTTCTGCTTGCCAAAACGAGAGGCTCTGGGTTCAAAATCCAGATGTG 1029
QY      1191 GTGTATGAGATATCAAGGGACAGAGTCAACAAAGGAGCCATAAATTTGATTGACAGC 1250
Db      1030 ACATACAAAGAACATACGTTGGAGACATCGCAACAGCAGCAGCAATTCATTAAGTCAAT 1089
QY      1251 ACAAACTTTCATGATGGAAGATTAATGAGAAATTAATTT 1294
Db      1090 GACGATGCCCCCTCCAAAAGATATAAGCTAGTATATCTTT 1133

```

```

RESULT 6
US-08-467-023-133
; Sequence 133, Application US/08467023
; Patent No. 6090386
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin U.;
; APPLICANT: Pollock, Joanne;
; APPLICANT: Bond, Julian F.;
; APPLICANT: Garman, Richard D;
; APPLICANT: Kuo, Mei-Chang;
; APPLICANT: Yeung, Siu-mei H.;
; APPLICANT: Brauer, Andrew;
; APPLICANT: Exley, Mark A.;
; APPLICANT: Powers, Steven P.
; TITLE OF INVENTION: Allergenic Proteins And Peptides From
; TITLE OF INVENTION: Japanese Cedar Pollen
; NUMBER OF SEQUENCES: 261
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
; STREET: 610 Lincoln St
; CITY: Waltham
; STATE: MA
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467.023
; FILING DATE: June 6, 1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/350 225
; FILING DATE: December 6, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38, 872
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 133:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 42..1586
; US-08-467-023-133

Query Match      13.9%; Score 228; DB 3; Length 1726;
Best Local Similarity 52.3%; Pred. No. 3,6e-43;
Matches 557; Conservative 0; Mismatches 495; Indels 12; Gaps 2;

QY      240 AATATTGACAAAGTTGATTAATAATGGGATTAAGATGATTAATGATTAAGTCTTGGAGCT 299
Db      180 AAGATTGAGCATCTCGTCATGATGCTATCAACATCTTCAATGTGGAAGATATGGCGCA 239
QY      300 AAGGTTATGGAATAACATATGATTAATTTGCAATTTTGACAGCATGGAATGAAGCATGT 359
Db      240 GTAGCGATGGAAGCATGATTTGCACTGAGGCAATTTTCACAGCATGGCAAGCTGCATGC 299
QY      360 TCATCTGAACACCTGTTCATTTTGTGCTCTTAAACAGAAATTTTGGATTCCTTAGAA 419
Db      300 AAAAACCCATCAGCAATGTGCTGTG---CCAGCGACAGAAATTTGTTGTTAAACAT 356
QY      420 ATCACTTTTCAGTCCATGACAGATCTTCTAATTTTCAGTAAAGATTTTGGATTCCTTAGAA 479
Db      357 CTGTTCTTCAATGGGCCATGTCMACTTACTTTTAAGTATGATGGATATATAGCT 416
QY      480 GCATCTGATTAATTTGACACTACAAAGATGAAGGCTTTGGATTCCTTTGATGATGT 539
Db      417 GCGTACCAAAATCCAGAGAGCTGGAAGAAATATGAAATGTGGACGTTTGTCAAACTT 476
QY      540 CAAATTTAGTGTGTTGAGAGAGAGAACTATCAATGGCAATGACAGATTTGTTGGCGCA 599
Db      477 ACAGTTTACTCTAATGGGTAAGAGTTAATTTGATGGCAAGAAAACATTTGTTGGCT 536
QY      600 AGTCTTGCAAAATTAATAATCACTGCCA-----TSCAGGATGACCAAGGCC 650
Db      537 GCCCATTAATTAATGGGTCAATGAGAGAAATTTGACAGATCGTGATAGACCAACAGCC 596
QY      651 TTAACCTTCTGGAATTTGCAAAATTTTGAAGTGAATCTTAAGAGTAAATGACACAA 710
Db      597 ATTAATTTGATTTTCCACGCGTCTGATTAATCCAAAGCATGAACTAATGAACGCTCC 656
QY      711 CAAATTCATCAAAATTTGATGCAATGCACTAATGTTGATGCTCAATTTGATGATCAAT 770
Db      657 GAATTTCAATTAGTTTGGGAATTTGAGGAGATTAATAATCATGGCATTTAGTATTCG 716
QY      771 GCTTCAGCAAGAGCCCAATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 830
Db      717 GACCGAAGACAGTCCATACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 776
QY      831 ATATCTGATCTATTTATGGAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 890
Db      777 TTACAAAAGAACACCATGAGAACAGAGGATGATGATGATGATGATGATGATGATGATGAT 836
QY      891 AATGTCAGGCGCAAAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 950
Db      837 AATATTGATTTGAGGATCTGATTTTGGCGTCCAGGCCATGGAATAGATATAGAAAGTCTT 896
QY      951 GGATCTGGAATTCAGAGCTTATGTGCTAATGTTACTGTAATGAAACCAAAATTAATC 1010
Db      897 GGGAGGGGAAAACCTAGACAGAGGTTTCACTGTCAGCTAATAATGGGGCTAAATTCATA 956
QY      1011 GGTGCCAAAATGAGTTAGATCAAGACTTGGCAGGAGAGATCTGGCAAGCTGCAAC 1070
Db      957 GACACACAAAATGATTAAGATCAAAACATGGAGGAGGTTTCAGGCAATGCGCAAGCCAT 1016
QY      1071 ATCAAAATTCGATGATGGAATGCAAGAGCTTAAGTATCCCATTAATTAAGACCAAAAC 1130
Db      1017 AATATTATGAGAAATGTAATGATTAATTTGGAGAACCCCATTAATTAATCAATTC 1076

```





Patent No. 5759829  
GENERAL INFORMATION:  
APPLICANT: SHEWMAKER, C.  
APPLICANT: KRIDL, J.  
APPLICANT: HIATT, W.  
APPLICANT: KNAUF, V.  
TITLE OF INVENTION: ANTI-SENSE REGULATION OF GENE  
TITLE OF INVENTION: EXPRESSION IN PLANT CELLS  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Filth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.1  
SOFTWARE: Microsoft Word 5.1 (a)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,213  
FILING DATE: 5-JUNE-95  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/750,505  
FILING DATE: 27-AUG-91  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/240,408  
FILING DATE: 30-AUG-88  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/920,574  
FILING DATE: 17-OCT-86  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/845,676  
FILING DATE: 28-MAR-86  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
NAME: Carl J. Schwedler  
REGISTRATION NUMBER: 36,924  
REFERENCE/DOCKET NUMBER: CGNE 26-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-08-463-213-1

[illegible]

Oy	442	GATCTTCTATTTCAGTAAAGATTTTGGATCC	473
Db	188	GATCTTCTATTTCAGTAAAGATTTTGGATCC	219
RESULT 9			
5453566-1			
; Patent No. 5453566			
; APPLICANT: SHEPMAKER, CHRISTINE K.; RIDL, JEAN C.; HIATT,			
; WILLIAM R.; KNUFF, VIC			
; TITLE OF INVENTION: ANTISENSE REGULATION OF GENE EXPRESSION			
; IN PLANT/CELLS			
; NUMBER OF SEQUENCES: 2			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/750,505			
; FILING DATE: 27-AUG-1991			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 240,408			
; FILING DATE: 30-AUG-1988			
; APPLICATION NUMBER: 920,574			
; FILING DATE: 17-OCT-1986			
; APPLICATION NUMBER: 845,676			
; FILING DATE: 28-MAR-1986			
; SEQ ID NO.:1:			
; LENGTH: 219			
5453566-1			

Query Match	13.0%	Score 212	DB 6	Length 219
Best Local Similarity	100.0%	Pred. No. 8	4e-40	
Matches 212	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	262	ATGGGATTAAAGTGATTATATGCTACTTACCTTTGAGCTTAAGGGGTGATGGAAAAATATATG	321	
Db	8	ATGGGATTAAAGTGATTATATGCTACTTACCTTTGAGCTTAAGGGGTGATGGAAAAATATATG	67	
QY	322	ATTAATATTCGACTTTGAGCAGACGATGGATGGAAGCATGTTCACTCTGACAACCTGTTCAAT	381	
Db	68	ATTAATATTCGACTTTGAGCAGACGATGGATGGAAGCATGTTCACTCTGACAACCTGTTCAAT	127	
QY	382	TTGTGGTTCCTAAAAACAAGAAATTATCTTCTCAGCAAAATCACTCTTTGAGGTCCATGCA	441	
Db	128	TTGTGGTTCCTAAAAACAAGAAATTATCTTCTCAGCAAAATCACTCTTTGAGGTCCATGCA	187	
QY	442	GATCTTCTATTTCCAGTAAGAATTTTGGATCC	473	
Db	188	GATCTTCTATTTCCAGTAAGAATTTTGGATCC	219	

```

10 RESULT 10
11 US-08-941-532-7
12 / Sequence 7, Application US/08941532
13 / Patent No. 6096946
14 /
15 / GENERAL INFORMATION:
16 /
17 / APPLICANT: ROBERTS, Jeremy Alan
18 / APPLICANT: COUPE, Simon Allan
19 / APPLICANT: JENKINS, Elizabeth Sarah
20 /
21 / TITLE OF INVENTION: CONTROL OF POD DEHISCENCE
22 /
23 / NUMBER OF SEQUENCES: 8
24 /
25 / CORRESPONDENCE ADDRESS:
26 / ADDRESSEE: Sterne, Kessler, Goldstein & Fox
27 / STREET: 1100 New York Avenue
28 /
29 / CITY: Washington
30 /
31 / STATE: D.C.
32 /
33 / COUNTRY: U.S.A.
34 /
35 / ZIP: 20005
36 /
37 / COMPUTER READABLE FORM:
38 /
39 / MEDIUM TYPE: Floppy disk
40 /
41 / COMPUTER: IBM PC compatible
42 / OPERATING SYSTEM: PC-DOS/MS-DOS
43 / SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
44 /
45 / CURRENT APPLICATION DATA:
46 /
47 / APPLICATION NUMBER: US/08/941,532
48 /
49 / FILING DATE: 30-SEP-1997
50 /

```

	Query Match	4.5%	Score 73;	DB 3;	Length 164;
	Best Local Similarity	67.3%;	Pred. No. 5.6e-08;		
	Matches 103;	Conservative 0;	Mismatches 50;	Indels 0;	Gaps 0;
QY	786	CCAAATATCTGATGGAGTCCATGTAATCAATATCTCAATATCTGATCTACTT	845		
Db	10	CCGAAATGCGAGCGGATTCATATCTGTTGCTATAAAACATTGCAATCTCCAAATTCAGAC	69		
QY	846	ATTGGACACAGGTGATGATTGTAATTTCAAATGTTCTTGAGATCTCAAAATGTGACGCCACA	905		
Db	70	ATTGGACACAGGTGATGATTGCAATATCCATTTAGAGTGAATGCAAAATGTTCAATCAAT	129		
QY	906	AATATTACTTGTGGTCCAGGTCAATGATTAAGT	938		
Db	130	GATTTAACCTTGGCGCCCGGCTACAGGCGCTAAGT	162		

? FILING DATE: 26-AUG-1991  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: BENT, Stephen A.  
 ? REGISTRATION NUMBER: 29,768  
 ? REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (703)836-9300  
 ? TELEFAX: (703)683-4109  
 ? TELEX: 899149  
 ? INFORMATION FOR SEQ ID NO: 14:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 7218 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 ? IMMEDIATE SOURCE:  
 ? CLONE: pTZ9C-F15  
 ? JS-08-232-463-14

```

Query Match      3.6%; Score 58.4; DB 1; Length 7219;
Best Local Similarity 5.0%; Pred. No. 0.00045;
Matches 20; Conservative 222; Mismatches 158; Indels 0; Gaps 0;

QY 930 GGTAATAAGTATGGAAGCTTAGATCTGCAAAATTCACAAGCTTATGTCTAATGTTACT 989
Db 1452 GATAGAAAGATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395
QY 990 GTAAATGAGACCCAAATTTATCGGTCCGAAAATGAGATTAGATCAAGACTTGGCAGGA 1049
Db 1392 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1333
QY 1050 GCAATCTGCAACACTGCAACATCAAAATTTCTGATATGGAATGCAAGCCTTAAGTAT 1109
Db 1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273
QY 1110 CCATATATTATAGACCAAACTATGTGATCGAGTGAACCATGATACCAAGCTTTCA 1169
Db 1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213
QY 1170 GCAGTTCAAGTGAAGAAAATGTGTGTATGAAATATCAAGGCAACAAGTGAACAAAGTG 1229
Db 1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153
QY 1230 GCCATATAAATTGATGTCAGACCAAACTTCCATGTGAAGAAATTATGAGAAATATA 1289
Db 1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1053
QY 1290 AATTGATAGGGGAAAGTGAAGAAACATCAGAGCGTTACGT 1329
Db 1092 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATGCAAGCTCCCT 1053

RESULT 12
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHITRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.

```

US-09-790-988-1

Query Match 3.2%; Score 53; DB 4; Length 640681;  
Best Local Similarity 51.0%; Pred. No. 0.035;  
Matches 125; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 1391 TTGAGAGATGAGAGCTCTTTGTTATATTTATATATCTATAGATCTTCATATATA 1450  
Db 325145 TTCTATATATATTTTATATATTTTAAATATTTCTTAAATATATTTTAAATGCTGA 325204  
QY 1451 GCAGATATGATATATCAATTAACAAATCTATATCTATGATTAATATATATTA 1510  
Db 325205 TTATATAAATAATATATATATATATTTTATATATATATATATATATATATAC 325264  
QY 1511 TATGTACGATGATGAGTTTATATAGCTACTATGTTCTATTTCTAGTCAAAAGTT 1570  
Db 325265 TATATAATTTTATATATTTTAAATATATATATATATATATATATATATATTTT 325324  
QY 1571 TGACGATGCTACTTTTATATGTCACAAATATATATATATATATATATATATAT 1630  
Db 325325 TTCTTTTATTTTATATATTTTAAATTTACTATATATATATATATATATATAT 325384  
QY 1631 AAAA 1635  
Db 325385 ATAA 325389

RESULT 13  
US-09-107-532A-2998  
Sequence 2998, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESS: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: &lt;Unknown&gt;

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-8277

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 2998:

SEQUENCE CHARACTERISTICS:

LENGTH: 1356 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1356  
SEQUENCE DESCRIPTION: SEQ ID NO: 2998  
US-09-107-532A-2998

Query Match 3.2%; Score 52.4; DB 4; Length 1356;  
Best Local Similarity 47.9%; Pred. No. 0.0061;  
Matches 183; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

QY 519 TGAATGCTTTGATGATGTTCAAAATTTAGTTTGGAGAGAGAGACTATCAATGCG 578  
Db 283 TGTATCTATGCTCAAAATGATGAGAACATCTCTGCTCACTGGTTTGGATGATGAGA 342  
QY 579 AATGACAGATGATGAGTGGCCCAAGTTCTTGCAAAATTAATTAATCACTGCCATGAGGAG 638  
Db 343 AATGAGAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 402  
QY 639 GCACCAACGGCTTACCTTCTGGAATGCAAAATTTGAAATGAAATATCTTAAAGAT 698  
Db 403 CCTAAATG--ATGAGTTTCCATATACGTGCTCATGAAATTAATGAAATTAATG 459  
QY 699 AAAATGACACCAATTCATATCAATTTGATGATGATGATGATGATGATGATGATGAT 758  
Db 460 ATCCAGTCACCAAGTTGAGCAGATCAACCCGATTTCTTGCAATATGACAGCTTGATAC 519  
QY 759 TTGATGATCAATGCTTCAACCAAGAGCCCAATTAATGATGATGATGATGATGATGAT 818  
Db 520 TTGACGATCTTGAACCCAGCGGATTCACCAATTAATGACAGGATTTGATGATGATG 579  
QY 819 CAATATATCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 878  
Db 580 AAAATGCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639  
QY 879 TCTGATCTCAATATGTCAGG 900  
Db 640 GCAGGAGCTGAAGATGATGATG 661

RESULT 14

US-09-790-988-1/C

Sequence 1, Application US/09790988

Patent No. 6632935

GENERAL INFORMATION:

APPLICANT: SHIGENOBU, SHUJI

APPLICANT: WATANABE, HIDEKI

APPLICANT: HATTORI, MASAHIRA

APPLICANT: SAKAKI, YOSHIYUKI

TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS

FILE REFERENCE: 081356/0159

CURRENT APPLICATION NUMBER: US/09/790,988

PRIOR FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: JP2000-04-07

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 640681

TYPE: DNA

ORGANISM: Buchnera sp.

US-09-790-988-1

Query Match 3.2%; Score 52; DB 4; Length 640681;  
Best Local Similarity 50.7%; Pred. No. 0.059;  
Matches 154; Conservative 0; Mismatches 145; Indels 5; Gaps 1;

QY 1333 AAAATGTCATTTTAAACAATGCTGAACATGTTACACCACTGCACCTTCACTAGAAATTT 1392  
Db 609613 AAAGTAGCAATTTCAATATACGATGATCTTTAGATTCAGTCCAGTCAATTTCAATATCT 609554  
QY 1393 CAGAGATGAGAGCTCTTTGTTATATTTTATATATATCTATAGATCTTCAATATATAGC 1452

Db 609553 ATCCAGATTAAATTTTATTTATGATTTTCATATTTTTCANAAATTAATAATTTT 609494  
QY 1453 AGATATGATATATACCAATTAACCAATCTATATCTATGATGAATTAATTAATA 1512  
Db 609493 TATTACATTAACCTGATTTTAATA-----AGATAAAAAAGAAATTTTAAATTA 609439  
QY 1513 TGTAAGATGAGTTTAAATAGACTATGATTTCTATTTTCTGATGAAAAAGTTG 1572  
Db 609438 ATTAATGATTTTATTTATTTATTTCTATATTAATATGATTAATTAATAAAAAATTT 609379  
QY 1573 ACGATGACTTTTAAATGACAAAAATTAATAAATGTTATTTATATGAAAAAAA 1632  
Db 609378 TTGATTAAGAAATATTTATTTATGAGAAAAATTTAAATATTTATATACATTAATAAATA 609319  
QY 1633 AAAA 1636  
Db 609318 AATA 609315

## RESULT 15

US-10-204-708-19/C  
; Sequence 19, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 19  
; LENGTH: 6866  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-19

Query Match 3.2%; Score 51.8; DB 4; Length 6866;  
Best Local Similarity 49.8%; Pred. No. 0.014;  
Matches 131; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1370 AACAAGCACTGACAGAAATTTGAGAGATGAAGCTCTTGTATATTAATTAATTTAT 1429  
Db 2406 AACAAGCACTGACAGAAATTTGAGAGATGAAGCTCTTGTATATTAATTAATTTAT 2347  
QY 1430 ACTATGATCTTCATATATAGATATGATATATGATATATGATATATGATATATGAT 1489  
Db 2346 AATAAAAACTAT 2287  
QY 1490 GATTTGAATTAATTAATTAATATATGATGATGATGATGATGATGATGATGATGAT 1549  
Db 2286 TTTTAAATATTTCTCTAAATATCCAAAAATTAATAAATTAATTAATTAATTAATTA 2227  
QY 1550 TCTATTTCTGAGTGAAGTTGAGATTTGATCTTTTATGATGATGATGATGATGATGAT 1609  
Db 2226 TCTCTAATAAATAAATAATTAACCAATAAAGCTATATATAAATAAATACTCAATATATA 2167  
QY 1610 GTTATTTATATGAAAAAAA 1632

Db 2166 TTATATATACTAAAAAACAA 2144

Search completed: July 27, 2004, 04:23:19  
Job time : 135 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 23:42:40 ; Search time 763 Seconds

(without alignments)  
10481.334 Million cell updates/sec

Title: US-10-018-604-1

Perfect score: 1636

Sequence: 1 aactcttttaacatagacaag.....atatgaaaaaaaaaaaaa 1636

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 3216467 seqs, 244149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1042.4	63.7	5822	13	US-09-924-197-1
2	420.2	25.7	1631	14	US-10-151-668-1
3	375	22.9	1359	17	US-10-437-963-24608
4	369	22.6	1182	17	US-10-437-963-24607
5	331.8	20.3	2207	13	US-09-783-130-26
6	281.2	17.2	1168	16	US-10-260-238-618
7	278.2	17.0	1335	17	US-10-437-963-28561
8	267.4	16.3	1308	13	US-10-425-114-30261
9	230	14.1	1809	17	US-10-437-963-24606
10	210.8	12.9	1333	13	US-10-424-599-61964
11	198	12.1	1219	13	US-10-425-114-10970
12	194.8	11.9	1479	13	US-10-424-599-141807
13	175	10.7	442	9	US-09-770-444-874
14	175	10.7	442	9	US-09-924-035A-369

15	171.2	10.5	1673	13	US-10-362-091-3	Sequence 3, Appli
16	169	10.3	1678	13	US-10-424-599-116781	Sequence 16781, A
17	168.6	10.3	1382	13	US-10-425-114-29325	Sequence 29325, A
18	164	10.0	870	13	US-10-424-599-59408	Sequence 59408, A
19	163.6	10.0	1984	13	US-10-424-599-92601	Sequence 92601, A
20	162	9.9	1631	13	US-10-425-114-9445	Sequence 9445, Ap
21	161	9.8	811	16	US-10-260-238-3398	Sequence 3398, Ap
22	154	9.4	596	17	US-10-021-323-9854	Sequence 9854, Ap
23	153.2	9.4	1472	13	US-10-424-599-95670	Sequence 95670, A
24	151.8	9.3	1185	9	US-09-938-842A-2520	Sequence 2520, Ap
25	151.8	9.3	1185	11	US-09-938-842A-2520	Sequence 2520, Ap
26	150.4	9.2	1272	17	US-10-437-963-14307	Sequence 14307, A
27	149.8	9.2	1067	13	US-10-425-114-16916	Sequence 16916, A
28	148.6	9.1	1374	17	US-10-437-963-43683	Sequence 43683, A
29	141.8	8.7	2012	13	US-10-424-599-33624	Sequence 33624, A
30	140.2	8.6	1432	13	US-10-425-114-7729	Sequence 7729, Ap
31	130.6	8.0	589	17	US-10-021-323-5922	Sequence 5922, Ap
32	130.6	8.0	746	9	US-09-770-149-31	Sequence 31, Appl
33	129.2	7.9	1248	16	US-10-260-238-498	Sequence 498, Appl
34	128.2	7.8	568	17	US-10-021-323-12284	Sequence 12284, A
35	126.6	7.7	2179	17	US-10-437-963-56617	Sequence 56617, A
36	125.6	7.7	565	13	US-10-424-599-83751	Sequence 83751, A
37	123	7.5	588	17	US-10-021-323-9731	Sequence 9731, Ap
38	121.6	7.4	1605	17	US-10-437-963-50404	Sequence 50404, A
39	120.2	7.3	602	17	US-10-021-323-9178	Sequence 9178, Ap
40	120.2	7.3	602	17	US-10-021-323-9178	Sequence 9178, Ap
41	118	7.2	1886	17	US-10-437-963-31774	Sequence 31774, A
42	117.8	7.2	1723	17	US-10-437-963-73025	Sequence 73025, A
43	117.6	7.2	588	17	US-10-021-323-9998	Sequence 9998, Ap
44	115	7.0	588	17	US-10-021-323-5864	Sequence 5864, Ap
45	115	7.0	588	17	US-10-021-323-9185	Sequence 9185, Ap

## ALIGNMENTS

RESULT 1  
US-09-924-197-1  
; Sequence 1, Application US/09924197  
; Publication No. US20030018993A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutterman, Neal  
; TITLE OF INVENTION: Improved Methods of Gene Silencing Using Inverted  
; TITLE OF INVENTION: Repeat Sequences  
; FILE REFERENCE: 012176-010810US  
; CURRENT APPLICATION NUMBER: US/09/924,197  
; CURRENT FILING DATE: 2001-08-07  
; PRIOR APPLICATION NUMBER: 60/225,508  
; PRIOR FILING DATE: 2000-08-15  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 5822  
; TYPE: DNA  
; ORGANISM: Agrobacterium tumefaciens  
US-09-924-197-1

Query Match 63.7%; Score 1042.4; DB 13; Length 5822;  
Best Local Similarity 99.4%; Pred. No. 3.5e-192;  
Matches 1046; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY	376	TTCAATTTGGTCTCTTAAACAGATTTATCTTCTCAAGCAATACCTTTTCAGGTC	435
DB	1379	TTAAACATGATGTTCTTAAACAGAAATTCCTTCAAGCAATACCTTTTCAGGTC	1438
QY	436	CATGAGATCTTCTTCAATTTAGATTTTGAATCTTGAACATCTAATAATT	495
DB	1439	CATGAGATCTTCTTCAATTTAGATTTTGAATCTTGAACATCTAATAATT	1498
QY	496	CAGACTAAAGATGAGAGCTTTGATGCTTTTGAATAGTTCATAATTAGTTGTTG	555
DB	1499	CAGACTAAAGATGAGAGCTTTGATGCTTTTGAATAGTTCATAATTAGTTGTTG	1558

```

QY 556 GAGAGAGAGAACTATCATATGCAATGACAGATATGAGCCAAATTTCTTGCAAAATA 615
DB 1559 GAGAGAGAGAACTATCATATGCAATGACAGATATGAGCCAAATTTCTTGCAAAATA 1618
QY 616 ATAAATACCTGCGATGAGGAGATGACCAACGGCTTAACCTCTGGAATTCGAAAAT 675
DB 1619 ATAAATACCTGCGATGAGGAGATGACCAACGGCTTAACCTCTGGAATTCGAAAAT 1678
QY 676 TGAAGAGATTAATCTAAAGATTAATAATGACAAACAAATTCATATCAATTTGAGTCAT 735
DB 1679 TGAAGAGATTAATCTAAAGATTAATAATGACAAACAAATTCATATCAATTTGAGTCAT 1738
QY 736 GCACTAAATGTTGATGCTTCAATTTGATGATCAATGCTTGACAAAGGCCAAATCTGTG 795
DB 1739 GCACTAAATGTTGATGCTTCAATTTGATGATCAATGCTTGACAAAGGCCAAATCTGTG 1798
QY 796 ATGAGATGCTATGATCAATTTGATGATCAATTTGATGATCAATTTGATGATCAATTTGATG 855
DB 1799 ATGAGATGCTATGATCAATTTGATGATCAATTTGATGATCAATTTGATGATCAATTTGATG 1858
QY 856 GTGATGATGATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 915
DB 1859 GTGATGATGATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1918
QY 916 GTGATGATGATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 975
DB 1919 GTGATGATGATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1978
QY 976 TGTCTAATGTTACTGTAATTAATGACCAAAATTAATGAGCCGAAATGAGATGATGATGATGATG 1035
DB 1979 TGTCTAATGTTACTGTAATTAATGACCAAAATTAATGAGCCGAAATGAGATGATGATGATGATG 2038
QY 1036 AGACTTGGCAGGAGATCTGAGACAGCTAGCAACATCAATTTCTGAATGTGAAATGTC 1095
DB 2039 AGACTTGGCAGGAGATCTGAGACAGCTAGCAACATCAATTTCTGAATGTGAAATGTC 2098
QY 1096 AAGAGCTTAAGTATCCATTAATTAATGACCAAAATTAATGAGCCGAAATGAGATGATGATGATG 1155
DB 2099 AAGAGCTTAAGTATCCATTAATTAATGACCAAAATTAATGAGCCGAAATGAGATGATGATGATG 2158
QY 1156 TACCAACATTTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1215
DB 2159 TACCAACATTTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2218
QY 1216 GTGCAACAAAGGTGCGCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1275
DB 2219 GTGCAACAAAGGTGCGCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 2278
QY 1276 TAATGAGATTAATTAATTAATTAATGAGGAGAAATGAGAAACCATGAGGCTTACGTGCAAAA 1335
DB 2279 TAATGAGATTAATTAATTAATTAATGAGGAGAAATGAGAAACCATGAGGCTTACGTGCAAAA 2338
QY 1336 ATGTGATTTTAAACAATGCTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1395
DB 2339 ATGTGATTTTAAACAATGCTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2398
QY 1396 AGGATGAGGCTCTTTTGTATTAATTTTAAATTT 1427
DB 2399 AGGATGAGGCTCTTTTGTATTAATTTTAAATTT 2430

```

## RESULT 2

```

US-10-151-668-1
; Sequence 1, Application US/10151668
; Publication No. US20020184660A1
; GENERAL INFORMATION:
; APPLICANT: ULVSKOV, Peter
; APPLICANT: CHILD, Robin
; APPLICANT: VAN ONCKELIN, Henri
; APPLICANT: PRINSEN, Els
; APPLICANT: BORKHARDT, Bernard
; APPLICANT: SANDER, Lilli

```

```

; APPLICANT: PETERSEN, Morten
; APPLICANT: BUNDEARD POUlsen, Gert
; APPLICANT: BOTTERMAN, Johan
; TITLE OF INVENTION: Seed Shattering
; FILE REFERENCE: 2121-0138P
; CURRENT APPLICATION NUMBER: US/10/151,668
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US/09/051,239
; PRIOR FILING DATE: 1998-09-28
; PRIOR APPLICATION NUMBER: PCT/EP96/04313
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: EP 95 402241.4
; PRIOR FILING DATE: 1995-10-06
; PRIOR APPLICATION NUMBER: EP 95 203328.0
; PRIOR FILING DATE: 1995-12-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; LOCATION: 95-163 = region encoding the presumed
; LOCATION: (95)..(1393)
; OTHER INFORMATION: Location 1229-1245 = region of the endo-PG cDNA
; OTHER INFORMATION: complementary to oligonucleotide PG5
; OTHER INFORMATION: complementary to oligonucleotide PG3
; OTHER INFORMATION: complementary to oligonucleotide PG2
; OTHER INFORMATION: complementary to oligonucleotide PG1
; NAME/KEY: CDS
; LOCATION: (95)..(1393)
; OTHER INFORMATION: Location 1229-1245 = region of the endo-PG cDNA
; OTHER INFORMATION: complementary to oligonucleotide PG5
; OTHER INFORMATION: complementary to oligonucleotide PG3
; OTHER INFORMATION: complementary to oligonucleotide PG2
; OTHER INFORMATION: complementary to oligonucleotide PG1
; NAME/KEY: unsure
; LOCATION: (1439)
; OTHER INFORMATION: Strain cv. Topaz.
; OTHER INFORMATION: n = a, c, g, t, any, other, unknown, or other
; US-10-151-668-1

```

```

Query Match 25.7%; Score 420.2; DB 14; Length 1631;
Best Local Similarity 63.6%; Pred. No. 1.4e-71;
Matches 717; Conservative 0; Mismatches 383; Indels 27; Gaps 4;

```

```

QY 277 TTAATGTAATGCTTGTGAGCTTAAGGTGATGAGAAACATATGATATATTCATTTG 336
DB 300 TTAGGTGTTGGAACCTTGGAGGCAAGAGATGAGAAACCGATGATCTCAGGCTTTCA 359
QY 337 AGCAGAGATGGAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 396
DB 360 AGAAGATGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 419
QY 397 ACNAGATTAATCTTCAAGCAATCACTTTTCAAGTCAATGATGATGATGATGATGATGATGATG 456
DB 420 GAAAGCTTAATCTTCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 479
QY 457 TTAAGATTTTGGATGCTTGAAGCACTAGTAAATTTTCAAGTCACTTCAAGTCACTTCAAG 513
DB 480 TCAAGTCTTGAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 539
QY 514 GGGTTGAGTGTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 567
DB 540 ACCACTGCTTATTTTGAAGAGCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATG 599
QY 568 CTATCAATGCAATGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 627
DB 600 TTGTTGATGCAACGGAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 659
QY 628 CATGAGGATGACCAACGAGCTTAACTTGAATGCAAAATTTGAAAGTGAATA 687
DB 660 CATGACAAAGGCGCAAGGCTTAACTTGAATGCAAAATTTGAAAGTGAATA 719

```



```
QY 688 ATCTAAGAGTAAATGACACAAATTCATATCAAAATTTGAGTCATGCACTAATGTTG 747
|||
Db 720 ATCTGAGAGTGAAGAAATGACACAGATTCAGATTTGCATGAGAAATGCAACAATGTTG 779
QY 748 TAGCTTCAAAATTTGATGATCAATGCTTCAGCAAGAGCCCAAAATATCTGATGAGATCCATG 807
|||
Db 780 GCGTTAAGAAATGTTAAGATCACTGCTCTGCGCATGATGCCAACAACGATGTAATTCATA 839
QY 808 TATCAAAATCTCAATATATTCAAAATATCTGATATCTATATTAATGGAACAGTGATGTTGA 867
|||
Db 840 TCGTTGCTACTAAAAACATTCGAAATCTCCAAATTCAGACATTTGGACAGTGATATTTGTA 899
QY 868 TTTCAATTTGTTTCGGAATTCCAAAATGTCAGGCCACAATATTAATTTGTTGTCAGATC 927
|||
Db 900 TATCCATTTGAGAGATGATCGCAAAATGTTCAAAATCAATGATTAATTAATTTGCGGCCGCTC 959
QY 928 ATGATTAATTAATTTGGAAGCTTAGGATCTGGAATTCGAAATTCGAAAGCTTAATGTTA 987
|||
Db 960 ATGGGATCAGCATTTGGAAGCTTGGGGATGACAAATTCGAAAGCTTAATGTTAATTCGGGAATG 1019
QY 988 CTGTAATTAAGAACCAAAATTAATGCTGCCGAAAATGAGTTAGGATCAAGACTTGGCAGG 1047
|||
Db 1020 ATGATGATGCTGCTACGCTCTGTAAGATGACAAATGAGTAAGATCAAGATTTACAGAG 1079
QY 1048 GAGGATCTGACAGCTAGACACATCAATTTCTGAATGTTGAAATGCAAGAGCTTAAGT 1107
|||
Db 1080 GAGGATCTGACAGCTAGACACATTAATTTCCAAAATTCGATGATGATATGTCAGA 1139
QY 1108 ATCCCAATTAATTAATGACCAAAATCATGATGATGAGTGAACATGATATCAACAGTTT 1167
|||
Db 1140 ATCCGATCATTAATGACCAAGCAACTACTGCGA---CAAGACAAATGGAACACACAGAT 1196
QY 1168 CAGCAGTTCAAGTGAAGAAATGATGATGAGAAATATCAAGGCAACAATGCAACAAG 1227
|||
Db 1197 CTGCGGTTCAAGTGAAGAAATGATGATGAGAAATATCAAGGCAACAATGCAACAAG 1256
QY 1228 TGGCCATTAATTTGATTTGACACAAATCTTCATGATGAGAAATTAATGAGAGATA 1287
|||
Db 1257 TGGCGTAATGTTTAATGATGATGAAATATCCATGCAAGGTAATGTTGCTTGAAGATG 1316
QY 1288 TAAATTTAGTGGGGAAGTGAAGAAACATCAGAGGCTACGTCAGAAAATGTCATTTTA 1347
|||
Db 1317 TGAACATCAAAAGAGGAA-----AACTCTTGGGAAAATGTCATGTTTA 1361
QY 1348 ACAATGCTGAACATGTTTACACCACTGCACTTCACTAGAAATTTCA 1394
|||
Db 1362 AGGATTAAGGCACTGTTCTCCTAATGCCCCCTTAATTAAGCTCA 1408
RESULT 3
US-10-437-963-24608
; Sequence 24608, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Bouhazuk, Andrew A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24608
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Oryza sativa
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_29576C.1
US-10-437-963-24608
Query Match 22.9%; Score 375; DB 17; Length 1359;
Best Local Similarity 60.0%; Pred. No. 7,66-63;
Matches 646; Conservative 0; Mismatches 425; Indels 6; Gaps 1;
QY 221 AAATATTGAAGCAACAATATATTTGACAGGTTGATTAATAATGAGATTAAGTCATTA 280
|||
Db 102 AACTAGTGAAGATGATAGTCTCAGAGGACAGTCTCTTGAAGTCAAAAGAGTGATGGA 161
QY 281 TGTACTAGCTTTGAGACCTAAGGCTGATGGAAGAAACATATATATATTTGATGAGCA 340
|||
Db 162 TGTGCGCAACATGCGCTTACGAGATGACACATGATGACAGAGGATTTGGCAAA 221
QY 341 AGCATGAATGAAGCATGTTCAATCTAGAACACCTGTTCAATTTGTGTTCTTAACAA 400
|||
Db 222 GGCATGGGCTGACAGCTTGTCTCTTTCGCAACCTTCATTTGTTCTCATCCAAAGGCGA 281
QY 401 GAATTAATCTCTCAGCAATCACTTTCAAGTCCATGCAAGATCTTCTATTTCACTAAA 460
|||
Db 282 GAGATACCTCAACAGCAATTAACATCTGTCATGCAATGCAAAATCAAGCATCACTTCAT 341
QY 461 GATTTTGGATCCCTTGAAGCATCTAGTAATTAATTTGACAT-----ACAAAGTAGAG 514
|||
Db 342 GATAGAGGGTACGTTGCTGCTCTCCAAAGAGTCAATTTGAGCAAGAAATTAATAG 401
QY 515 GCTTTGATGCTTTTGAATGTTCAAAATTAATTTGTTGAGAGAGAGAACTATCAA 574
|||
Db 402 GCACGTGATTAATGTTCAATGTTGATGAGTGTGCTTACTGTGCTGCTGTTGAGATCTGCA 461
QY 575 TGGCAATGACAGATATGTTGAGCAAGTTTTCGAAATTAATTAATCACTGCCATGAG 634
|||
Db 462 TGAAGACGCAATATTTGTTGCGAAATTTCTCCAAACCAATGCAAAATCTTCATGAC 521
QY 635 GATGACCAACGCGCTTAACCTTCTGGAATTCGAAATTTGAAGTGAATTAATCTAAA 694
|||
Db 522 TGAAGCTCAACGCGCTTGAATTTTACTCTCTTTCAAACTTAAGTGAAGAACTTGAA 581
QY 695 GAGTAAAAATGCAACAAATTCATCAAAATTTGAATGATGATCACTAATTTGATGCTTC 754
|||
Db 582 GCTACTAAACAGCAACAAATTCACATGTCAGTTGAGATTTGACATGATTTGAGATCTC 641
QY 755 AAATTTGATGATCAATGCTTCAAGCAAGAGCCCAATTAATGATGAGTCCATGATCAA 814
|||
Db 642 TAGCTGAATCAATCAACGACGACGACGACGACGACGACGACGACGACGACGACGACG 701
QY 815 TACTCAATATTAATCAATATCTGATCTATTAATTTGGAACAGGATGATTTGATTCAT 874
|||
Db 702 AGGTAAAAATGTAACAAGTACAGGCTGCTTAATTAACAAGCCGGGATGATGCTGAT 761
QY 875 TGTTCGTGATCTCAAAATGTCAGGCCCAAAATATTAATTTGCTGTCAGGTCATGAT 934
|||
Db 762 TGAAGATGAACTGAGAACTTACATGTCAGAAACATGATGAGACCGGGAACGCGCAT 821
QY 935 AAGTATTTGAAGCTTGAAGATCTGGAATTCAGAAATTCAGAAATTTGATGATTTGAT 994
|||
Db 822 CAGCATCGGAGCTTGAAGGATCAACAATCTTAAGGTCATGTCACAAATGTCACCGTCA 881
QY 995 TGAAGCCAAATTAATCGTGCAGAAATGGAATGAGATCAAGATCTGAGAGGAGATC 1054
|||
Db 882 CACCGTCAGGCTATATGAGCAACAACGAGAGCTGCAATCAAGATGAGAGGCGGTTG 941
QY 1055 TGAAGAGCTAGCAACATCAATTTCTGAATGGAATGCAAGAGCTTAATGATCCAT 1114
|||
Db 942 GGGCTACGGAAGAACATGCTGTTCCAGAACTGATCAAGAAATTTGTTGAACCAAT 1001
QY 1115 AATTATAGCAAAATCTTTGATGCTGAGTTGAACATGATATATCAACAGTTTACAGAT 1174
|||
Db 1002 CATTAATCAACGAACATCTGATTTGCTACACCTTCGCAAGAACAGACATCTGCGGT 1061
QY 1175 TCAAGTGAATAATGATGATGAGATATCAAGGCAACAAGTGAACAAAGTGGCAT 1234
```

Db 1062 GCAAGTAGCAATGTGCTCTCAAGAACATCAGGGGCAAGGCTCTCAAGAGGCTAT 1121  
 QY 1235 AAAATTTGATTGAGCAGCAAACTTTCCATGTGAGAGAAATTATATGAGAGATATATA 1291  
 Db 1122 CAATGCTGATGTCAGCAAAATGACTTGTCTCAAGAAATTAACCTTGAAGAGCGTCMA 1178

## RESULT 4

US-10-437-963-24607  
 ; Sequence 24607, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53321)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 24607  
 ; LENGTH: 1182  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_29575C.1  
 US-10-437-963-24607

Query Match 22.6% Score 369; DB 17; Length 1182;

Best Local Similarity 60.1%; Pred. No. 1e-61;

Matches 634; Conservative 0; Mismatches 415; Indels 6; Gaps 1;

QY 263 TGGGATTAAAGTGAATTAATGTAAGTCTTGGAGCTAAGGGTGATGAGAAAACATATGA 322  
 Db 78 TGGCAGTAATGTGTTACACATACAGAGCTACGGGGCTCATGGAGCGACGATGATGA 137  
 QY 323 TAATATTCATTTGAGCAAGCATGGAATGAGCATGTTCATCTTGAACACCTGTCANT 382  
 Db 138 CACCAAAAGCATGGAGATACATGGCTGCACTTGCCTTGCAGAAACCTGCAGTTT 197  
 QY 383 TGTGTTCTTAAACACAGAAATATCTCTCAAGCAATCACCTTTGAGTCCATGAG 442  
 Db 198 GCTATCCCAAGGCAAGAAATACCTGATCAAGACACACACACTGTCTGGTCCATGCA 257  
 QY 443 ATCTTCTATTCAGTAAGATTTTGGATCCTTAGAAGCATGTAAATTTCAAGCT- 501  
 Db 258 ATCAAGCATCTCATTTGATGTGAAGGTATTTGGTGGCTTCTCCAGAGAGCTCAGACT 317  
 QY 502 -----ACAAAGTGAAGGCTTTGATGCTTTTGAATGATGTTCAAAATTTAGTGTGG 556  
 Db 318 GAGCAAGGAGACCATTAAGCACTGATTTGATCAAGTGTGTCATCTGCTTACTGTGAC 377  
 QY 557 AGAGAGAGAACTATCAATGAGCAATGAGCAAGATGTGGCAAGTTCTTGGAAAATATA 616  
 Db 378 TGGTGTGGGACCATAGATGAGAAATGCAAGATTTGGTGGCAAAATTCATGCAAAACCA 437  
 QY 617 TAAATCATGCTCATGAGGAGTCAACCAACGGCTTTACCTTCTGGAATGCAAAATTT 676  
 Db 438 CTCGAAGCTTCATGACAGAAAGCTCCAAACGGCACTGATCTTACTCTGCAAGATCT 497  
 QY 677 GAAAGTGAATTAATCTAAAGATTAATGCAACAAATTCATTAATTTGAGTCAG 736  
 Db 498 GAAGGTAGATTAATTAAGGTGTGAACGACCAAGAAATTCATTAATTTGAGTCAG 557  
 QY 737 CACTATGTTGATGCTTCAAAATTTGATGATCAATGCTTCAAGCAAGAGCCCAATATCTGA 796

Db 558 CACCATGTGATGTGTCTGCTGCTGATCAGACACAGAAACAGCCCAACTGA 617  
 QY 797 TGGAGTCCATGTAATCAATATCTCAATATATTTCAATATCTGATATTTATTTGAGACAG 856  
 Db 618 TGAATTCATATACACGCGACAGAGATGTTGAGTACAGATGATGATCAAGACCG 677  
 QY 857 TGATGATTTGATTTCAATTTTCTGATCTCAAAATGTGAGGCCCAAAATATTACTTG 916  
 Db 678 GGATGACTGATGATCAATGACAGAGACGGAACCGAATCTGATGTCAGAAACATGTGTG 737  
 QY 917 TGGTCAAGTCAATGATTAATGATTTGAGACCTTGAATCTGAAATTCAGAACTTATGT 976  
 Db 738 TGAACCGGACACGCGCATGACATCGGTAGCTGGATGATCATATTTCTGAACTCATGT 797  
 QY 977 GTCTAATGTTACTGTAATGAAGCCAAATTAATGCGTCCGCAAAATGAGTTAGATCA 1036  
 Db 798 CAACATGTCACTGTCACATATGTCAGTGTGTATGACAGACCCAGAGAGTCCATCA 857  
 QY 1037 GACTTGGCAGGAGATCTGACAAAGCTTACACATCAATTTCTGAATGTGAAATGCA 1096  
 Db 858 GACATGCGAGGAGAGAAAGGCTTACAGCAAGAACATGCTGTTCCAGAAACATGTCATGGA 917  
 QY 1097 AGAGTTAATGATCCCAATTAATTAAGCAAACTATTTGATGATGATGATGATGAT 1156  
 Db 918 CAATGTCTGAAACCCATCATCATGACACCAAACTACTGCACTCTTACACCTGCA 977  
 QY 1157 ACAACAGTTTTCACAGCTTCAAGTGAAGAAATGTGATGATGATGATGATGATGATGAT 1216  
 Db 978 GCAACAGAAATCTGATGAGAGGATGAGCAATGTGTCTCAAGAACATCAGGGGCAAG 1037  
 QY 1217 TGCAACAAAGGTGCGCATTAATTTGATTTGAGACACAACTTTCATGTGAAAGATAT 1276  
 Db 1038 TGATTCAGAGAGGACCATCATGCTGATTTGAGAGAGAGTGTACTCTTCCATGATAC 1097  
 QY 1277 AATGAGAAATTAATTTAGTAGGAGAAAGTGA 1311  
 Db 1098 CTGAGAAATGTCATCTCACTGTCACAGGAGAA 1132

## RESULT 5

US-09-782-130-26  
 ; Sequence 26, Application US/09782130  
 ; Publication No. US20040055038A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KNAUF, VIC C.  
 ; APPLICANT: KNAUF, VIC C.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATED TRANSCRIPTION  
 ; FILE REFERENCE: 16518.052  
 ; CURRENT APPLICATION NUMBER: US/09/782,130  
 ; CURRENT FILING DATE: 2001-02-12  
 ; PRIOR APPLICATION NUMBER: US 09/232,861  
 ; PRIOR FILING DATE: 1999-01-15  
 ; PRIOR APPLICATION NUMBER: US 08/812,665  
 ; PRIOR FILING DATE: 1997-03-07  
 ; PRIOR APPLICATION NUMBER: US 08/484,941  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: US 08/105,852  
 ; PRIOR FILING DATE: 1993-08-10  
 ; PRIOR APPLICATION NUMBER: US 07/526,123  
 ; PRIOR FILING DATE: 1990-05-21  
 ; PRIOR APPLICATION NUMBER: US 07/267,685  
 ; PRIOR FILING DATE: 1988-11-02  
 ; PRIOR APPLICATION NUMBER: US 06/592,605  
 ; PRIOR FILING DATE: 1985-01-17  
 ; PRIOR APPLICATION NUMBER: US 07/582,241  
 ; PRIOR FILING DATE: 1990-09-14  
 ; PRIOR APPLICATION NUMBER: US 07/188,361  
 ; PRIOR FILING DATE: 1988-04-29  
 ; PRIOR APPLICATION NUMBER: US 07/168,190  
 ; PRIOR FILING DATE: 1986-03-15  
 ; Remaining seq Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 35

```

; SOFTWARE PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-782-130-26

Query Match
Best Local Similarity 95.3%; Score 331.8; DB 13; Length 2207;
Matches 342; Conservative 0; Mismatches 17; Indels 0; Gaps 0

QY      1  AATCTTTTCAATAGCAAGTTTAAAAACCATACCATATACATATATCATGTGTTATCC 60
      |||
DB      1430 AATCTTTTCAATAGCAAGTTTAAAAACCATACCATATACATATATCATGTGTTATCC 1485
      |||

QY      61  AAAGAAATGTAATTCCTTCCTTCATATATATTTTGGCTTCATCATTTCACTTTAGAA 120
      |||
DB      1490 AAAGAAATGTAATTCCTTCCTTCATATATATTTTGGCTTCATCATTTCACTTTAGAA 1545
      |||

QY      121 GCATATGTTATGATGACATTTTATTCACCAAGTTTATGATGATATCTTGACCAAGAT 180
      |||
DB      1550 GCATATGTTATGATGACATTTTATTCACCAAGTTTATGATGATATCTTGACCAAGAT 1609
      |||

QY      181 TTGCATCATATTTTCAAGTTATCTTTCTTATTTGAGCAAAAATTTGAAAGCAACATA 240
      |||
DB      1610 TTGCATCATATTTTCAAGTTATCTTTCTTATTTGAGCAAAAATTTGAAAGCAACATA 1669
      |||

QY      241 ATATTGACAAGGTTCATTAATAATGGATTTAAAGTCATTAATGACTTAAGCTTTGAGCTA 300
      |||
DB      1670 ATATTGACAAGGTTCATTAATAATGGATTTAAAGTCATTAATGACTTAAGCTTTGAGCTA 1729
      |||

QY      301 AGGGTGATGAAAAACATATGTAATATTGCAATTTGACGACGATGGAATGACGATGT 359
      |||
DB      1730 AGGGTGATGAAAAACATATGTAATATTGTAAGATTTTAATATGGAATATATTGT 1788
      |||

RESULT 6
US-10-260-238-618
; Sequence 618, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyuki
; APPLICANT: Krepes, Joel
; APPLICANT: Proxart, Nicholas
; APPLICANT: Riche, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 618
; LENGTH: 1168
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (8)..(8)
; OTHER INFORMATION: n = any nucleotide
; NAME/KEY: N_region

```

/	LOCATION:	(46)..(46)
/	OTHER INFORMATION:	n = any nucleotide
/	FEATURE:	
/	NAME/KEY:	N_region
/	LOCATION:	(49)..(49)
/	OTHER INFORMATION:	n = any nucleotide
/	FEATURE:	
/	NAME/KEY:	N_region
/	LOCATION:	(102)..(102)
/	OTHER INFORMATION:	n = any nucleotide
/	FEATURE:	
/	NAME/KEY:	N_region
/	LOCATION:	(190)..(190)
/	OTHER INFORMATION:	n = any nucleotide
/	US-10-260-238-618	

  

Query Match	17.2%;	Score 281.2;	DB 16;	Length 1168;
Best Local Similarity	55.1%;	Pred. No. 1.2e+44;		
Matches 613;	Conservative 0;	Mismatches 487;	Indels 12;	Gaps 3

  

QY	273	GTGATTATATCTTACCTTTGGAGCTTAAGGCTGATGTAAGAAACAATCATGATAATTTGCA	332
Db	35	GTTTTTAGCTTGANANGTTACGGGGGAACGGCGAAGCAGAACGCCGACACACGGCGGC	94
QY	333	TTTGACACGATGGAATAGAAGATGTTTCATTAGAACCTGTTCAATTGTGGTCTC	392
Db	95	CTTCCCANAGCCCTTGAAACCCGCCCTGGGGCTGGGGCGCGCGCGCTGCTGCTGCCG	154
QY	393	AAAAACAAGATTTATCTTCAAGCAAAATCACCTTTCAAGTCCATGACGATCTTCTAT	452
Db	155	GAGAGAGAGAGCTACTGCTCCGGGAAGTCACCTTTCGGGCCCTTGCAATCCACATC	214
QY	453	TGATTAAGATTTTGGATCTTAAAGATCTAGTAAATTTGACTACAAAG-----	507
Db	215	AAGCTCATGTGAAGGGGACGCTGGTGCGTGCCGCACTGCAATGCAATCGAAGCAGAC	274
QY	508	-ATGAAGCTTTGGATTTGCTTTTATATGTTTCAAATTTGATTGTTGAAGAGAGA	566
Db	275	AACAGAGAGTACTGATGCTGCTCGCGGCGTCAAGCGGCTGCGCGTGGCGGGCGGC	334
QY	567	ACTATCATAGGAATGAGCAGATATGATGGGCCAAGTTTGGCAATTAATTAATCACTG	626
Db	335	ACCATTCGACGGCAGCGGAGGGGGTGGGAGAACTCTCCCAAGATCAACAGGGCACTC	394
QY	627	CCATCGACGAGTGCACCAAGCGCTTAACTTTGGAATTSCAAAAATTSAAAGTAAT	686
Db	395	CCATCGAAGGAGGCTCCCAAGCGCGTGAGTTTCCACAGTGCACACCTTAGCGTGAT	454
QY	687	AATCTAAAGATTAATAATGCACAAATTCATTAATTAATTTGAATGACATTAATGTT	746
Db	455	GCTCGAAGATCGTGAACAGCCACACAGATCCATATGTCAGTGGAGATTGACCGGGGTG	514
QY	747	GTAGCTTAATTTGATGATCAATGCTTCAGCAAAAGGCCCAATTAATGATGAGTCAAT	806
Db	515	GAGCTGGCCCACTGTCAATCTTGGCGCCGGGCGAGGCCCAACAGATGGATCCAC	574
QY	807	GTATCAATTAATCAATATATCAATATCTGATCTGATTTTGGAGCAGGATGATGTT	866
Db	575	ATCACCCACACAGAAAGCTCAGTACGACATCGACCATTTAAGACAGGGGAGTACGT	634
QY	867	ATTTCATATTGTTTGTGATCTCAAAATGTGACGGCCAATAATTAATTTGATGTCAGAT	926
Db	635	GTGTCGATCGAGATGGGACCCACGTTTACAGTCAAGAGACTGATGTGGGCCCGGG	694
QY	927	CATGTAATTAATTTGAGAGCTTAGATCTGGAATTCAGAAAGTTATGTGTTAATGTT	986
Db	695	CATGGATTAACATTTGAGAGCTTAGAGATGACAACTCCAGAGTGAAGGTCTGACATC	754
QY	987	ACTGTAATGAGCCAAATTTATGCGGCCGAAATTTGATGATGATGATGATGATG	1046
Db	755	TTATCAACACCGTGACCTCTTATGAGACCAACCAATGAGAGCTGATCAAGATGCGAG	814
QY	1047	GGAGAGATTCGACAGCTGACCAATCAATTTTGTGATGTTGAAATGCAAGCGTTAAG	1106

Db 815 GAGGGAGTGGATACGCGCAAGGATATCGATTTCCAGAACATGCTCATGAACAGTGTCCAG 874  
QY 1107 TATCCCATATTTATGACCAAACTATTTGATGAGTTGAACATG--TATACACAG 1163  
Db 875 AACCCATATATCTTTGACCAAACTACTGCACTCAAGAAATGCGAGACAGAG 934  
QY 1164 TTTTCAGAGTTCAAGTGAATAATGTGTGTATGAGAAATTCAGAGGCAAGTGCACA 1223  
Db 935 GATCAGCAGTGGAGATTAACCAATGTGTCTTCAGAACATCGAGGACACAAATTTCC 994  
QY 1224 AAGTGGCCATAAATTTGATGTCAGACAACTTTCCATGTGAGAAATTAATGAG 1283  
Db 995 AAGATGTCATCTACTCTGAAGTGCAGCAAGACTACCAATGCTATGACATTTCTTACG 1054  
QY 1284 AATATAATTTAGTAGGGGAAAGTGAA--ACCATCAGAGGCTACGTGCAAAATGTC 1340  
Db 1055 GACATCACTCGAAATGTGTGATGACATGTGTCCACAGGAAGTACTTGCAGAAATGCA 1114  
QY 1341 CATTTTACAAATGCTGAACATGTTACCCACA 1372  
Db 1115 AATGAGGAAATCTGGAAACAGTTGTTCCACA 1146

RESULT 7  
US-10-437-963-28561  
; Sequence 28561, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Mu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 28561  
; LENGTH: 1335  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_33148C.1  
US-10-437-963-28561

Query Match 17.0%; Score 278.2; DB 17; Length 1335;  
Best Local Similarity 55.3%; Pred. No. 4.8e-44;  
Matches 607; Conservative 0; Mismatches 478; Indels 12; Gaps 3;

QY 288 AGCTTTGAGCTAAGGGTGTATGAAAAACATATGATTAATGATTCATTTGAGAGAGCATGG 347  
Db 217 AGCTACGGCGCGACGCGACGCGAGCGCGACGACGCGCGCGCTGCGAGGCGCTGG 276  
QY 348 AATGAAGCATTTTCATCTTGAACACCTGTTCAATTTGTGTTCTTAATAAACAAGATTAT 407  
Db 277 AGCGCGCGCTGCGCTCGCGCGCGCGCGCGCGCTGCTGCTCCCGAGAGCGAGACTAC 336  
QY 408 CTTCAGCAAAATACCTTTTCAAGTGCATGCAATCTTTCTATTTTCACTAAAGATTTT 467  
Db 337 CTGCTCCGCGAAGTCACTCTCCGCGCGCTGCGAATTCACATCAAGCTCATGTGAAG 396  
QY 468 GGATCTTTGAAGCATCTAGTAAATTTCAAGCTACAAAGATAGA-----AGGCTTGG 521  
Db 397 GGAGCGCTGTGGCGCTGCGCGAGCATGTGAACTGGAACGAGAGCAAGAGATCTGG 456  
QY 522 ATTGCTTTGATAGTGTCAAAATTTAGTTGTGAGAGAGGAAGTATCAATGCAAT 581

Db 457 ATCGTGTCCGCGCGCTGCAAGGCGCTGCGCGCGCGCGCGCACATCGACGGCAC 516  
QY 582 GCAAGATATGTGGCCAAATTTCTTGCAAAATTAATTAATCACTGCGCATCAGAGATGA 641  
Db 517 GCGGAGGGGTGGGAGAACTCTGCAAGATGACAGGCGCATCCCATCGAAAGGAGCT 576  
QY 642 CCAAGCGCTTAACTTTCTGGAATTTGCAAAATTTGAAAGTGAATATCTAAAGATGAA 701  
Db 577 CCAAGCGCTGAGTTTCCACAGTGTGCAACCTGAGCTGATGTGTGAAGATGCTG 636  
QY 702 AATGCAACAATAATTCATATCAATTTGATCATGCACTATGTGTAGCTTCAATTTG 761  
Db 637 AACAGCCAGAGATCCACATGTGATGAGAGATTTGACCGGGGTGAGCTGCGCCACTG 696  
QY 762 ATGATCATGCTTCAGCAAGAGCCCAATATCTGATGAGTCCATGTATCAATATCTCA 821  
Db 697 TCAATCTCTGCGCGCGCGCGAGGCCCAACAGATGCAATCCATCCACAGCAAG 756  
QY 822 TATATTCAAATATCTGATCTATTTATTTGAAACAGTGTATGTATTTCAATTTCT 881  
Db 757 AAGTCCAGTCAAGCGACTGCAACCATTAAGACAGGGATGACTGTGTGATGAGAT 816  
QY 882 GATCTCAAAATGTCAGCGCGCAAAATTAATCTTGCTCAGGTCAATGTATTAATTT 941  
Db 817 GGAAGCCAGGTTTACAGTTCAGAGTGTGTGTGGCGCGCGCGCGCATGGATTAAGATT 876  
QY 942 GGAAGCTTAGATCTGGAATTTCAAGCTTATGTGTCTAATTTACTGTAAATGAAGCC 1001  
Db 877 GGAAGCTTAGATCTGGAATTTCAAGCTTATGTGTCTAATTTACTGTAAATGAAGCC 1001  
QY 1002 AAAATTTATGCTGCG 1061  
Db 937 CACCTTATGCGCACACCAATGAGCTGATCAAGACATGCGAGGAGGAGTGTATC 996  
QY 1062 GCTGCAACATCAATTTCTGATGTGGAATGCAAGAGCTTAAATGATCCATTAATTA 1121  
Db 997 GCGAAGTATCTGATTTCAAGATGATGTCATGAACATGTCCAGAACCAATATCAAT 1056  
QY 1122 GACCAAACTATTTGATGAGTGAACCATG--TATACACAGTTCACAGCTTCAA 1178  
Db 1057 GACCAAACTATTTGATGAGTGAACCATG--TATACACAGTTCACAGCTTCAA 1178  
QY 1179 GTGAAAAATGTTGTATGAGATTTCAAGGCGCAAGTGCACAAAGTGGCCATTA 1238  
Db 1117 ATTAGCATGTGTGCTTCAAGAACTCCAGAGCAACAAATTTCAAGATGCTCACT 1176  
QY 1239 TTTGATGCGACAGCAACTTTCCATGTGAGAGATTTATGAGAAATTAATTTAGTA 1298  
Db 1177 CTGAATCTCAGCAAGACTACCATGTCTATGACATTTCTTACAGAGATCAACTGGA 1236  
QY 1299 GGGGAAAGTGAAG--ACCATCAGAGGCTACGTGCAAAATGTCATTTTAACAATGCT 1355  
Db 1237 ATGTGAGATGACAAATGTGTGCGCAGGAAGTACTTCCGAAATGCAAAATGGAATCT 1296  
QY 1356 GAACATGTTACCCACA 1372  
Db 1297 GGAACAGTTGTTCCACA 1313

RESULT 8  
US-10-425-114-30261  
; Sequence 30261, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jinsong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B

```
/ CURRENT APPLICATION NUMBER: US/10/425,114
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 30261
/ LENGTH: 1308
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: UC-ZMFLB73002C03_FLI
US-10-425-114-30261
```

```
Query Match      16.3%; Score 267.4; DB 13; Length 1308;
Best Local Similarity 58.7%; Pred. No. 5.9e-42;
Matches 463; Conservative 0; Mismatches 326; Indels 0; Gaps 0;
```

```
QY 500 CTACAAAGATGAAAGGCTTGGATTGCTTTTGATAGTGTTCAAATTTAGTTGTTGAGG 559
DB 64 CGAAGATGACGAGGACATGATCGTGTCCGAGATGACAGAGCTCACCGCACAAG 123
QY 560 AGGAGGAATGATCAATGGCAATGACAAAGTATGTTGCGCAAGTTCTTGCAAAATAA 619
DB 124 CGGTGGCGCATGATGACGAGCGGCAAAATGTTGCTCATCTCGTCAAGATCAACAA 183
QY 620 ATCACTGCCATGACGAGGATGACCAACGCGCTTAACCTTGGAATGCAAAATTTGAA 679
DB 184 GGCCTCCCTTTGGAAGAGGCTCCGAGGCTCTGTCACTTCCACTGCGTCACTGAA 243
QY 680 AGTGAATATCTAAGAGTAAATGACAAACAATTCATATCAAAATTTAGTCATGAC 729
DB 244 AGTGGAGATCTGAATGTAATGTAAGCAAGCAAGATGATCATGTGATGAGATTTGTC 303
QY 740 TATGTTGATGCTTCAATTTGATGATCAATGCTTACGAAAGAGCCCAATATCTGATG 799
DB 304 AAATGCGCTGCGCGCGGTTGTCATCAACGCGCTGCGACCGCTTAACCTGACGG 363
QY 800 AGTCATGATCAATATCAATATATATCAATATGATGATATATTTAGTAAACAGTGA 859
DB 364 CATCCATCATCGCGAGCAAGAAATGTACCGCTCAACAGTCAAGATCAAGCAAGGGA 423
QY 860 TGATTTATTTCAATTTGTTCTGGATCTCAAAATGTCAGGCGCAAAATTTACTTTGG 919
DB 424 CCACTGCAATGATGATGAGAAAGCGGACTCAACCTTCATGTTCCAAAGTTAAGTGTG 483
QY 920 TCCAGGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 979
DB 484 TCCAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 543
QY 980 TATGTTACTGTAATGAAGCCAAATTTANCGGTGCGAAATGAGTTAGATCAAGAC 1039
DB 544 CGGCAATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 603
QY 1040 TTGGCAAGGAGATCTGCAAGACTGACAAATTCAAATTTGAAATGTAAGTGAAGA 1099
DB 604 GTACCAAGGAGGAGGAGGATGAGCCCAAGGACATCAAGTTCCAAATTTGTTATG 663
QY 1100 GGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1159
DB 664 GGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
QY 1160 AAGATTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1219
DB 724 ACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 783
QY 1220 AAGCAAGTGGCATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1279
DB 784 TACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 843
QY 1280 GAGGATAT 1288
DB 844 GAGGATAT 852
```

RESULT 9

```
US-10-437-963-24606
/ Sequence 24606, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Zhou, Yinhua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 24606
/ LENGTH: 1809
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_29574C.1
US-10-437-963-24606
```

```
Query Match      14.1%; Score 230; DB 17; Length 1809;
Best Local Similarity 60.0%; Pred. No. 1.3e-34;
Matches 383; Conservative 0; Mismatches 255; Indels 0; Gaps 0;
```

```
QY 657 TTCTGAATTTGCAAAATTTGAAAGTGAATATCTAAGAGTAAATGCAACAATTT 716
DB 142 TGCAGAGAGACTTTATTTCTGAAGTGAAGACCTGAGTGTGAACAGCCAGCAATC 201
QY 717 CATATCAATTTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 776
DB 202 CAGATTTCAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 261
QY 777 GCAAGAGCCCAATTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 836
DB 262 GAAATCTGCCCCAAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 321
QY 837 GATATATTTATTTGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 896
DB 322 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 381
QY 897 GAGGCAACAATTTACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 956
DB 382 CATGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 441
QY 957 GGAATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016
DB 442 CATATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 501
QY 1017 GAAATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1076
DB 502 AAGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 561
QY 1077 TTTGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136
DB 562 TTTGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 621
QY 1137 GATGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196
DB 622 GATGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 681
QY 1197 GAGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1256
DB 682 AAGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 741
QY 1257 TTTGATTTGAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1294
DB 742 GTGCTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 779
```

RESULT 10  
US-10-424-599-61964  
; Sequence 61964, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 61964  
; LENGTH: 1333  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRI3847\_26969C.1  
US-10-424-599-61964

Query Match 12.9%; Score 210.8; DB 13; Length 1333;  
Best Local Similarity 54.0%; Pred. No. 5.9e-31;  
Matches 431; Conservative 0; Mismatches 367; Indels 0; Gaps 0;  
DB 512 AAGCTTTGATGCTTTTGTAGTGTTCATAAATTTAGTTGTGAGAGAGAGAACTAT 571  
19 ACGGTTTGGCTTATTTTCCAAATGACAAATTTTCCAAAGGTCGTGAGTTAT 78  
QY 572 CAATGGCAATGACAAAGTATGAGTGGCCCAAGTTCTTGCAAAATTAATACAGCCATG 631  
79 TGATGGCTCAGAAAGCAAAATGATGGCGCATCTTGCAAAAAGAACAAAGTCCATCTTG 138  
QY 632 CAGGATGACCAACCGGCTTAACCTTCTGGAATTCGAAAATTTGAAAGTGAATATCT 691  
139 CAAGGTGACCAACAGATTTACATTTGATGATCAAGTTCAATCCATTAAGGTGAAGCACT 198  
DB 692 AAAGAGTAAATGACAAACAAATTCATATCAAAATTTAGTCAGTACCTAATGTTGAC 751  
199 AACATCCAGATGACCAAGATGATCTTTTATCTATTCAGATGATGATCTGTTAGAT 258  
QY 752 TTCAATTTGATGATCAATGCTTTCAGCAAAAGCCCAATATCTGATGAGTCCATGATC 811  
259 TACTGGCGTGAAGTGTGACGACCTGAGACAGCCCAACACTGATGATGATCAATATG 318  
DB 812 AAATCTCAATATATCAATATCTGATCTATTAATTTGGAACAGGTGATGATGATTTTC 871  
319 TGAATCAACAAATGTCATATTCACAAAGCAAGCAAAATTTGGAACAGGTGATGATGATC 378  
QY 872 AATTGTTCTGATCTCAAAATGTCAGAGCCCAAAATATTAATTTGTTGTCAGGTGATG 931  
379 AATTGTCATGCTAGTCTAATATCAAAATGGAAGAAATTTATTTGAGCCAGCAATG 438  
DB 932 TATAAGTATGGAAGCTTATGATCTGGAATTCAGAACTTCAAGCTTATGCTCAATGTTACTG 991  
439 AATCAGCATTTGGAAGCTTATGGAAGCAAACTCAACAGGATGATGATCAAAAGTATTTT 498  
QY 992 AAATGAGCAAAATTAATGATGCTGCGGAAATGAGTTAGATCAAGACTTGGCAGGAG 1051  
499 GGAATCAGAGATTTCTTAGGAGATCAACAAAGGTCTCAAGATTAAGACTTTGGAGGAG 558  
DB 1052 ATCTGCAAGCTAGCAACATCAAAATTTCTGATGATGGAATGCAAGCTTATGATCTC 1111  
559 TTCTGATATGTTGAGGGGTGCTTTTCAGATGATGAGGGTGAAGAACTCGAGCGAT 618  
QY 1112 CATTAATTAAGCAAAATCTATTGATGATGAGTTGAACATGATATCAACAGTTTCAAG 1171  
619 CGGCTCGAGAGCAATTTTACTGATGATCTCCACCAATTTGGAATTAAGCATCAGC 678

QY 1172 AGTCAAGTGAATAATGTGTATATGAAATATCAAGGAGAGAGCAACAAAGGTG 1231  
DB 679 AGTGAGATTAAGCAGAGTATGATCAAGCAATTAAGTGCATCAATGCTGTAAGGC 738  
QY 1232 CATTAATTTGATTTGAGCAACAACTTCCATGCTGAAGATTAATGAGATTAAT 1291  
DB 739 CATTAATTTGATTTGAGCAACAACTTCCATGCTGAAGATTAATGAGATTAAT 798

QY 1292 TTTAGTAGGGGAAAGTGG 1309  
DB 799 CTTAGAGAAACAAAGATGG 816

RESULT 11  
US-10-425-114-10970  
; Sequence 10970, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; NUMBER OF SEQ ID NOS: 2003-04-28  
; SEQ ID NO 10970  
; LENGTH: 1219  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 70092481\_F11  
US-10-425-114-10970

Query Match 12.1%; Score 198; DB 13; Length 1219;  
Best Local Similarity 55.1%; Pred. No. 1.7e-28;  
Matches 387; Conservative 0; Mismatches 315; Indels 0; Gaps 0;  
DB 608 CAAATTAATTAATCACTGCGATGAGGATGACCAACAGGCTTAACCTTCTGAAATG 667  
1 CAAAAGAACAGTCAATCTTTCGAAAGTGACCAACAGCATTTACATGATGATCAAG 60  
QY 668 CAAAATTTGAAAGTGAATTAATCTAAGAGTAAATGACAAACAAATTCATATCAAT 727  
DB 61 TTCAATCCATAGGAGTGAAGGATCAACATCCAGATGACCAAGATGATGATGAT 120  
QY 728 TGATGATGACTAATGTTGATGCTTCAAAATTTGATGATGATGATGATGATGATG 787  
DB 121 ATCAGATGATGATGCTGTTAGAAATTAATCTGCGTGAAGTGCAGACCTGAGACG 180  
QY 788 AAATCTGATGAGTCAATGATCAAAATTAATCAATATTAATTAATGATGATGAT 847  
DB 181 AAACCTGATGAAATTCATATTAATGATCAACAAATGATCAATATTCAGACGAAAT 240  
QY 848 TGAACAGGTGATGATGATGATTTCAATTTCTGATCTCAAAATGTCAGGCGCAAA 907  
DB 241 TGAACAGGAGTATGATGATGATCAATTTCAATGCTCAATATCAAAATGATGAG 300  
QY 908 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 967  
DB 301 AATTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 968 AGCTTATGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027  
DB 361 AGCATATGTCAGAAAGTATTTTGAATACAGAGTCTTTAGGAGATCAACAGGTCT 420  
QY 1028 TAGATCAAGACTTGGCAGGAGATCTGACAAAGCTAGCAACATCAAAATTTCTGAAT 1087  
DB 421 CAGATTAAGACTTGGCAGGAGGTCTGATATGTTGAGGGGTGCTTTTCAAGATGT 480

QY 1088 GGAAATGCAAGCTTAAGTATCCCATTAATATAGACCAAACTATTGTGATGCAATTGA 1147  
| | | | |  
Db 481 GAGGTGGAAATATATCCAAACCCATTATTAAGACCAATTATCTGTGATTCCTCAAC 540  
| | | | |  
QY 1148 ACCATGATATCAACAGTTTTCAGAGCTTCAAGTAAATGCTGTATGAGATATCA 1207  
| | | | |  
Db 541 CAATTGGAAATATGAGCATTCAGATGAGATGAGATGAGATGATGATGATGATGATG 600  
| | | | |  
QY 1208 GGGCAACAATGTCACAAAGGTGGCATTAATTAATTTGATTTGATGAGCAACAATTTCCATGTA 1267  
| | | | |  
Db 601 TGGCACTACATCAATAGTCTAAGGCTTAATTAATTTGATGATGATGATGATGATGATG 660  
| | | | |  
QY 1268 AGGAATTAATATGAGATTAATTAATTTGATGAGGAAATGCG 1309  
| | | | |  
Db 661 CAAACTAGTCTTACGACAGTGACTTAAGAGAAAACAAGATGG 702  
| | | | |

## RESULT 12

US-10-424-599-141807  
; Sequence 141807, Application US/10424599  
; Publication No. US2004031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223) B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 141807  
; LENGTH: 1479  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99063C.1  
US-10-424-599-141807

Query Match 11.9%; Score 194.8; DB 13; Length 1479;  
Best Local Similarity 53.5%; Pred. No. 8e-28; Indels 33; Gaps 4;

Matches 518; Conservative 0; Mismatches 417; Indels 33; Gaps 4;  
QY 273 GTGATTAATGATCTTAGCTTTGAGCTAAGGCTGATGAAAAACATATGATTAATTTGA 332  
| | | | |  
Db 286 GTTTTCGATGTGAGATCTTTGGGGCTTTGAGATGGCTGTGATGATGACACAGTGA 345  
| | | | |  
QY 333 TTTGAGCAAGCATGGAATGAAAGCATGTTTATCTAAGAACCTGTTCAATTTGTGTTCT 392  
| | | | |  
Db 346 TTCAAGCAAGCATGGAAGCAAGCATGT--GCTGTGACTCTGGGATTTGTTCTGCTCA 402  
| | | | |  
QY 393 AAAAACAAGATTAATCTTCTCAAGCAAAATCACCTTTTCAGGTCATGCGATCTTCTAT 452  
| | | | |  
Db 403 GAAAACCTACAGTTTAAATATCACTTCACTATTTTTCAGGTCATGCAAGCAAGATGG 462  
| | | | |  
QY 453 TCAGTAAAGATTTTGGATCCTT-----AGAAGCATCTAGTAAATTTTCAAGTAC 503  
| | | | |  
Db 463 GTATCCAGTGAAGTGAACACTAATGCAACAGATGACAAATTCGTGGCCAAAGCA 522  
| | | | |  
QY 504 AAAAGTAAAGGCTTTGATGCTTTGATGATGTTTCAAAATTTAGTTTGGAGAGAGA 563  
| | | | |  
Db 523 GATGCGCCAAATCAATGCTGTGATTTTATGCACTTGACCAATATGACTTTAATGTGACA 582  
| | | | |  
QY 564 GGAATTCATATGCAATGCAACAGTATGTTGG-----CCAGTTCT 605  
| | | | |  
Db 583 GGAACCATTTGAAGGAAATGAGAACAAATGTTGGGATCTTCCCTGCAACCTTACAGGGGT 642  
| | | | |  
QY 606 TGCATAATTAATTAATCACTGCAATGCAAGGATGCAACGAGCTTCACTTGTGAT 665  
| | | | |  
Db 643 CCCAATGGAATAAATTTGTCAAGGCGCATGTGTAGTCTGTATGATACGGTTCTTCAATG 702  
| | | | |

QY 666 TGCATAATTTGAAGTGAATTAATCTAAGATGAATGAACAAATTCATATCAAA 725  
| | | | |  
Db 703 AGCTCCAAATTTGAAGTGAAGGGGTTGAAAATTCATAAACAAGTCTCTGATTCACATGTA 762  
| | | | |  
QY 726 TTTGATGATGCACTAATGTTTGAAGCTTCAAAATTTGATGATCAATGCTTCAGCAAGAGC 785  
| | | | |  
Db 763 TTCAATGCTGCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 822  
| | | | |  
QY 786 CCAATATCTGATGAGTCCATGATATCAATCAATCAATCAATCAATCAATCAATCAATCA 845  
| | | | |  
Db 823 CCCAACCCGATGGAATCAATGTAAGAAATTTCCAAATTTGGGATTAATTAATTTCAATG 882  
| | | | |  
QY 846 ATTGAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 905  
| | | | |  
Db 883 ATTAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 942  
| | | | |  
QY 906 AATATTACTTTGTGCTCAGGTCAATGATTAATGATGATGATGATGATGATGATGATGATG 965  
| | | | |  
Db 943 GGTTTAATCTTTGTGCTCAGGTCAATGATTAATGATGATGATGATGATGATGATGATGATG 1002  
| | | | |  
QY 966 GAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1025  
| | | | |  
Db 1003 CAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1062  
| | | | |  
QY 1026 GTTGAATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1085  
| | | | |  
Db 1063 CTCAGATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1122  
| | | | |  
QY 1086 GTGGAATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1145  
| | | | |  
Db 1123 ATCCAAATGAGAACTTTGAAACTGATCATCATATGACCAATATGATGATGATGATGATG 1182  
| | | | |  
QY 1146 GAACATGATTAACAAGTTTTCAGAGTTCAGTAAATGAAGCCAAATTTATCGGTATGATGATG 1205  
| | | | |  
Db 1183 GA--ATGCTAACAAGCACTTCACTGATGATGATGATGATGATGATGATGATGATGATG 1239  
| | | | |  
QY 1206 AAGGCGAC 1213  
| | | | |  
Db 1240 AAGGGAAC 1247  
| | | | |

## RESULT 13

US-09-770-444-874/C  
; Sequence 874, Application US/09770444  
; Patent No. US20020023280A1  
; GENERAL INFORMATION:  
; APPLICANT: Goriach, Jörn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Krieker, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hubban, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; FILE REFERENCE: 2027 (PABA-016PRV)  
; CURRENT APPLICATION NUMBER: US/09/770,444  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR FILING DATE: 60/178,502  
; NUMBER OF SEQ ID NOS: 999

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 874  
LENGTH: 442  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-874

Query Match 10.7%; Score 175; DB 9; Length 442;  
Best Local Similarity 65.0%; Pred. No. 3.1e-24;  
Matches 275; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

963 TCAGAAAGCTTATGCTTAATGTTACTGTAATGAAGCCAAATTTATCGTCCGAAAT 1022  
Db TCAGAAAGCTTATGCTTAATGTTACTGTAATGAAGCCAAATTTATCGTCCGAAAT 1022  
441 TCAGAAAGCTTATGCTTAATGTTACTGTAATGAAGCCAAATTTATCGTCCGAAAT 382  
QY 1023 GAGATTAGATCAAGACTTGGCGAGGAGATCTGGAACAGCTAGCAATCAATTTCTG 1082  
Db 381 GGTGTGAGAAATCAAGACTTGGCGAGGAGATCTGGAACAGCTAGCAATCAATTTCTG 322  
QY 1083 AATGTGAAATGCAAGCGTTAGTATCCATTAATTTAGCCAAACTATTGTGATCGA 1142  
Db 321 GACATCTAATGAAAAAGCTTACAAACCCCAATTAATCAACAGGACTACTGGATCGT 262  
QY 1143 GTTGAACCATGTATACAAAGTTTTCAGAGTTCAAGTGAATAATGTGTGTATGAGAT 1202  
Db 261 GTTGAAGCATGCCCCGAAAGAAATCCGCGGTACAGAGTGAATGTGTGTATGAGAT 202  
QY 1203 ATCAAGGGCAAGTGAACAAAGGTGGCCATTAATTTGATTGACAGCAAACTTTCCA 1262  
Db 201 ATACAAAGGACGAGCTCAAGACCCCATAGCTGTGAATTTGTATCAGCAAGAACATTTCA 142  
QY 1263 TGTGAAGAAATTAATAGGAAATATTAATTTAGTAGGGAAAGTGGAAAA--CCATCA 1319  
Db 141 TGTGAGAAATTAATAGGAAATATTAATTTAGTAGGGAAAGTGGAAAA--CCATCA 1319  
QY 1320 GAGGCTACGTGCAAAAATGTCATTTTAACAATGCTGAACATGTTACACCACTGCACT 1379  
Db 81 AAAGCTTCTTGCTCCATGTGAAGCTTGAGACCCGAGAAATGTTTCTCTTTGCACT 22  
QY 1380 TCA 1382  
Db 21 TGA 19

RESULT 14  
US-09-924-035A-369/c  
Sequence 369, Application US/09924035A  
Patent No. US20020142319A1  
GENERAL INFORMATION:  
APPLICANT: Grilach, Jm  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2011US  
CURRENT FILING DATE: 2000-08-11  
PRIOR FILING DATE: 1999-08-13  
NUMBER OF SEQ ID NOS: 900  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 369  
LENGTH: 442  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-924-035A-369

Query Match 10.7%; Score 175; DB 9; Length 442;  
Best Local Similarity 65.0%; Pred. No. 3.1e-24;  
Matches 275; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

963 TCAGAAAGCTTATGCTTAATGTTACTGTAATGAAGCCAAATTTATCGTCCGAAAT 1022  
Db 441 TCAGAAAGCTTATGCTTAATGTTACTGTAATGAAGCCAAATTTATCGTCCGAAAT 382

QY 1023 GAGATTAGATCAAGACTTGGCGAGGAGATCTGGAACAGCTAGCAATCAATTTCTG 1082  
Db 381 GGTGTGAGAAATCAAGACTTGGCGAGGAGATCTGGAACAGCTAGCAATCAATTTCTG 322  
QY 1083 AATGTGAAATGCAAGCGTTAGTATCCATTAATTTAGCCAAACTATTGTGATCGA 1142  
Db 321 GACATCTAATGAAAAAGCTTACAAACCCCAATTAATCAACAGGACTACTGGATCGT 262  
QY 1143 GTTGAACCATGTATACAAAGTTTTCAGAGTTCAAGTGAATAATGTGTGTATGAGAT 1202  
Db 261 GTTGAAGCATGCCCCGAAAGAAATCCGCGGTCAAGATGAGCAATGTGTGTATGAGAT 202  
QY 1203 ATCAAGGGCAAGTGAACAAAGGTGGCCATTAATTTGATTGACAGCAAACTTTCCA 1262  
Db 201 ATACAAAGGACGAGCTCAAGACCCCATAGCTGTGAATTTGTATGCAAGAACATTTCA 142  
QY 1263 TGTGAAGAAATTAATAGGAAATATTAATTTAGTAGGGAAAGTGGAAAA--CCATCA 1319  
Db 141 TGTGAGAAATTAATAGGAAATATTAATTTAGTAGGGAAAGTGGAAAA--CCATCA 1319  
QY 1320 GAGGCTACGTGCAAAAATGTCATTTTAACAATGCTGAACATGTTACACCACTGCACT 1379  
Db 81 AAAGCTTCTTGCTCCATGTGAAGCTTGAGACCCGAGAAATGTTTCTCTTTGCACT 22  
QY 1380 TCA 1382  
Db 21 TGA 19

RESULT 15  
US-10-362-091-3  
Sequence 3, Application US/10362091  
Publication No. US20040049809A1  
GENERAL INFORMATION:  
APPLICANT: Instituto de Ciencia Aplicada e Tecnologia (ICAT)  
TITLE OF INVENTION: Pear genes codifying for b-galactosidase, Pectin Methyl-esterase,  
FILE REFERENCE: none  
CURRENT FILING DATE: 2003-02-12  
PRIOR FILING DATE: 2000-08-22  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 1673  
TYPE: DNA  
ORGANISM: Pyrus communis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (112)..(1308)  
OTHER INFORMATION:  
US-10-362-091-3

Query Match 10.5%; Score 171.2; DB 13; Length 1673;  
Best Local Similarity 54.9%; Pred. No. 3.2e-23;  
Matches 360; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

QY 639 GCACCAAGGCGCTTAACCTTCTGAATGCAAAAATTTGAAGTAATCTAAAGCT 698  
Db 562 GAGGACACTACACTAGCTTTTCGAACCTCCAAACGTTGTGTGATGATTAATATCA 621  
QY 699 AAAAAAGCAACAATAATCAATCAATTTGATGATGATGATGATGATGATGATGAT 758  
Db 622 CTAAACAGCAAAATGTTCCACTTGTGCAACGCGCTGCCAAAATGTGAAATGCAAGGT 681  
QY 759 TTGATGATCAATGCTTTCAGCAAGAGCCCAATTAATGATGATGATGATGATGATGAT 818  
Db 682 GTCAAGGTTAAAGCGCGCGGCAACAGCCCAACCGATGATGATGATGATGATGATGAT 741  
QY 819 CAATATATTCAATATCTGATCTATTTATTTGAACAGTGAATGATTTTCAATGTT 878  
Db 742 TCTGAGATCACTTCTGACTTCCAAATTTTCAACCGGTGAGGACGTGTCTGATTTGGC 801



QY 879 CTGGATCTCAAAAATGTCAGAGGCCCAAAATATTACTTGTGGTCCAGGCTCATGTGTAAGT 938

Db 802 CCCGSCATCAACCAATTTGTGATTGAAACCGTCATGTGACCCGGCCAGGAATCGC 861

QY 939 ATTGGAGCTTAGATCTCGAAATTCAGAAGCTTATGTCTAAATGTTACTGTAATGAA 998

Db 862 ATTGGAGATTAGGGAAGAGCAACCAAGAAGCCGCTGTACAAAATGTTTACGTTAAACA 921

QY 999 GCCAAATTTATCGGTGCGCCGAAAATGAGAGTTAGATCAAGACTTGGCAGGAG--GATCT 1055

Db 922 GTTACATTCACCTGCTACTGAAAAACGGCCCTCAGAAATTAAGCTTGGGGGAGACCTYAGCCT 981

QY 1056 GGACACAGGTAGCAATCATCAAAATTTCTGATGAGTGAAGAAATGCAAGAGCTTAATATCCATA 1115

Db 982 GGATTTGCTTGAGAGCAATCTTTTCCACACTATGTGATGACCAACGTTCAAAAATTCACATC 1041

QY 1116 ATTATAGACCAAACTATTGTGATCGAGTTGAACCATGTATATCAACAGTTTTCAGCACT 1175

Db 1042 GTTATATGATCAAAATTACTGCCCTATATGACAAAGGTTGCCCTGGCGCAAGCTTCTGGAGTT 1101

QY 1176 CAAGTGAAAAATGTGTGTATGAGATATCAAGGCGACAGTGCACAAGAGTGGCCATA 1235

Db 1102 AAGGTACCGCATGTGACCGTATCAAGACATTCATGCTATCATGGGAGCGGAAGTGGCGGTG 1161

QY 1236 AAATTGTATGCAGACAAACCTTTCACATGTCAAGGAATTAATATGGAATATATAA 1291

Db 1162 AAATTCAGTTAGTTCATGTATATCTTTCGCAACGGGATCAACTGCAAGATGTGAA 1217

Search completed: July 27, 2004, 04:36:07  
Job time : 766 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 23:07:30 ; Search time 4442 Seconds

(without alignments)  
10998.318 Million cell updates/sec

Title: US-10-018-604-1

Perfect score: 1636  
Sequence: 1 aactctttccaatagacaag.....atatgaaaaaaaaaaaaa 1636

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_huv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rdc:\*  
26: em\_gss\_png:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	778.6	47.6	797	12	BM410796 EST585123
2	770	47.1	786	12	BM412817 EST587144
3	765.4	46.8	767	12	BM412528 EST586855
4	754	46.1	774	12	BM413004 EST587331

5	742	45.4	742	12	BM413282	BM413282	EST587609
6	738	45.1	728	12	BM413343	BM413343	EST587670
7	726.4	44.4	749	12	BM413075	BM413075	EST587402
8	729	43.3	729	10	AM222257	AM222257	EST299068
9	707.4	43.2	709	10	AM441939	AM441939	EST311335
10	694.4	42.4	686	10	AM442253	AM442253	EST311649
11	683.8	41.8	687	10	AM442335	AM442335	EST311731
12	674	41.2	703	12	BM536089	BM536089	EST589111
13	673.8	41.0	677	10	AM442241	AM442241	EST311637
14	670.2	41.0	675	12	BM408546	BM408546	EST582873
15	657.2	40.2	670	12	BM411547	BM411547	EST585874
16	657	40.2	672	10	AM222222	AM222222	EST299033
17	654.4	40.0	672	10	AM221785	AM221785	EST298956
18	643	39.3	643	10	AM223400	AM223400	EST300211
19	642.4	39.3	644	10	AM442052	AM442052	EST311448
20	639	39.1	639	10	AM442280	AM442280	EST311676
21	631	38.6	632	10	BE434493	BE434493	EST405571
22	630.4	38.5	632	12	BM410538	BM410538	EST584865
23	628.8	38.4	632	10	BE432845	BE432845	EST399470
24	628.4	38.4	630	10	BE431613	BE431613	EST336428
25	623	38.1	623	12	BM409252	BM409252	EST583579
26	617.4	37.7	619	12	BM413202	BM413202	EST587529
27	617	37.7	617	10	BE460938	BE460938	EST412357
28	604	36.9	604	10	BE433321	BE433321	EST398850
29	603	36.9	614	10	BE433180	BE433180	EST397019
30	602.4	36.8	604	10	BE461204	BE461204	EST412623
31	597.8	36.3	601	10	AM222835	AM222835	EST299646
32	595.4	36.4	598	10	AM223911	AM223911	EST300722
33	588	35.9	600	10	AM442198	AM442198	EST311594
34	583.4	35.7	585	10	BE436917	BE436917	EST408035
35	583.4	35.7	593	10	AM222726	AM222726	EST299537
36	580	35.5	600	10	AM441498	AM441498	EST310894
37	579	35.4	582	12	BM409146	BM409146	EST583473
38	574.4	35.1	576	10	AM222864	AM222864	EST299675
39	572	35.0	572	10	AM223468	AM223468	EST300279
40	570.4	34.9	572	10	AM222731	AM222731	EST299542
41	570.4	34.9	588	10	BE432533	BE432533	EST399062
42	567	34.7	567	10	AM223661	AM223661	EST300472
43	565	34.5	566	10	BE434462	BE434462	EST405540
44	563	34.4	563	10	AM441402	AM441402	EST310798
45	561.8	34.3	565	10	BE460910	BE460910	EST412329

#### ALIGNMENTS

RESULT 1  
LOCUS BM410796 797 bp mRNA linear EST 22-JAN-2002  
DEFINITION EST585123 tomato breaker fruit Lycopersicon esculentum cDNA clone  
C1EG5415 5' end, mRNA sequence.  
ACCESSION BM410796  
VERSION BM410796.1 GI:18262426  
KEYWORDS EST.

SOURCE  
ORGANISM Lycopersicon esculentum (tomato)

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 797)  
Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A.,  
Tsai, J., Bougri, O., Kirkness, E., Utecherack, T., Van Aken, S.,  
Romling, C.M., Fraser, C.M., Martin, G.B., Tankeley, S.D. and  
Giovannoni, J.

Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: COGI

TITLE  
JOURNAL  
COMMENT  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute

FEATURES Seq primer: T3.  
Location/Qualifiers  
Source 1..797  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEG54L15"  
/issue\_type="Pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/clone\_idb="tomato breaker fruit"  
/note="Vector: pBluescriptSKmCudapt; Site 1: EcoRI;  
Site 2: XhoI; supplier: Boyce Thompson Institute; fruit  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN

Query Match 47.6%; Score 778.6; DB 12; Length 797;  
Best Local Similarity 99.4%; Pred. No. 2.3e-120;  
Matches 792; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 359 TTCAATGAGACACCTGTTCAATTTGGTTCCTAAACAGAAATATCTTCTCAAGCA 418  
DB 1 TTCAATGAGACACCTGTTCAATTTGGTTCCTAAACAGAAATATCTTCTCAAGCA 60

QY 419 AATGACCTTTTCAGTCCATGACAGATCTTCTATTCTAGTAAGAATTTTGGATCCTTGA 478  
DB 61 AATGACCTTTTCAGTCCATGACAGATCTTCTATTCTAGTAAGAATTTTGGATCCTTGA 120

QY 479 AGCATCTAGTAAGATTTCAAGCTCAAGATAGAAAGGCTTTGGATGCTTTTGTAGTGT 538  
DB 121 AGCATCTAGTAAGATTTCAAGCTCAAGATAGAAAGGCTTTGGATGCTTTTGTAGTGT 180

QY 539 TCAAAATTTAGTTGTGGAGAGAGGAGAACTATCAATGGCAATGGAGCAAGTGGTGGCC 598  
DB 181 TCAAAATTTAGTTGTGGAGAGAGGAGAACTATCAATGGCAATGGAGCAAGTGGTGGCC 240

QY 599 AAGTCTTGGCAAAATTAATAATCACTGCCATGACGAGGATGACCAACGGCTTAACCTT 658  
DB 241 AAGTCTTGGCAAAATTAATAATCACTGCCATGACGAGGATGACCAACGGCTTAACCTT 300

QY 659 CTGGAATTTGCAAAATTTGAAAGTGAATATCTAAAGCTAAATATGACACAAATTTCA 718  
DB 301 CTGGAATTTGCAAAATTTGAAAGTGAATATCTAAAGCTAAATATGACACAAATTTCA 360

QY 719 TATCAAAATTTAGTCATGACACTAATGTTGTAGCTTCAAAATTTGATGATCAATGCTTCAGC 778  
DB 361 TATCAAAATTTAGTCATGACACTAATGTTGTAGCTTCAAAATTTGATGATCAATGCTTCAGC 420

QY 779 AAAGAGCCCAATATCTGATGAGTCCATGTATCAAAATATCTCAATATATCTCAATATCTGA 838  
DB 421 AAAGAGCCCAATATCTGATGAGTCCATGTATCAAAATATCTCAATATATCTCAATATCTGA 480

QY 839 TACTATTTATTTGGAACAGGTGATGATTTTCAATTTTCTTGGATCTCAAAATGTCGA 898  
DB 481 TACTATTTATTTGGAACAGGTGATGATTTTCAATTTTCTTGGATCTCAAAATGTCGA 540

QY 899 GGCCCAAAATTTCTTGTGGTCCAGTCAAGTGTAAATGATTTGGAAGCTTGAAGATCTGG 958  
DB 541 GGCCCAAAATTTCTTGTGGTCCAGTCAAGTGTAAATGATTTGGAAGCTTGAAGATCTGG 600

QY 959 AAATTCAGAGCTTATGTGTCTAATGTATCTGTAATGAGCCAAATTTATCGGTGCGCA 1018  
DB 601 AAATTCAGAGCTTATGTGTCTAATGTATCTGTAATGAGCCAAATTTATCGGTGCGCA 660

QY 1019 AAATGAGTTAGATCAAGACTTGGCAGGAGAGATCTGGAAGCTAGACATCAATTAATT 1078  
DB 661 AAATGAGTTAGATCAAGACTTGGCAGGAGAGATCTGGAAGCTAGACATCAATTAATT 720

QY 1079 TCTGATGTTGGAATGACAGAGCTTAAGTATCCCATATATTATAGACCAAACTATT-GTG 1137  
DB 721 TCTGATGTTGGAATGACAGAGCTTAAAGTATCCCATATATTATAGACCAAACTATTGTTG 780

QY 1138 ATCGAGTTGAACCATGT 1154  
DB 781 ATCGAGTTGAACCATGT 797

RESULT 2  
BM412817 786 bp mRNA linear EST 22-JAN-2002  
LOCUS EST587144 tomato breaker fruit Lycopersicon esculentum cDNA clone  
cLEG61A20 5' end, mRNA sequence.  
BM412817  
VERSION BM412817.1 GI:18264447  
KEYWORDS EST.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 786)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Uteerback,T., Van Aken,S.,  
Rommig,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and  
Giovannoni,J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.

FEATURES Location/Qualifiers  
source 1..786  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEG61A20"  
/issue\_type="Pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/clone\_idb="tomato breaker fruit"  
/note="Vector: pBluescriptSKmCudapt; Site 1: EcoRI;  
Site 2: XhoI; supplier: Boyce Thompson Institute; fruit  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN

Query Match 47.1%; Score 770; DB 12; Length 786;  
Best Local Similarity 98.7%; Pred. No. 6.3e-119;  
Matches 776; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 333 TTGAGCAGAGATGAGATGAGCATGTTCACTATAGAACACCTGTTCATTTGTGGTTCT 392  
DB 1 TTGAGCAGAGATGAGATGAGCATGTTCACTATAGAACACCTGTTCATTTGTGGTTCT 60

QY 393 AAAAAAGAAATATCTTCTCAAGCAATCACTTTTCAAGTCCATGACATCTTCTATT 452  
DB 61 AAAAAAGAAATATCTTCTCAAGCAATCACTTTTCAAGTCCATGACATCTTCTATT 120

QY 453 TCAGTAAGATTTTGGATCTCTTAAGACATCTAGTAATTTTCAAGCTACCAAGTATAG 512  
DB 121 TCAGTAAGATTTTGGATCTCTTAAGACATCTAGTAATTTTCAAGCTACCAAGTATAG 180

QY 513 AGGCTTGGATTGCTTTTGTATAGTGTCAAAATTTAGTGTGGAGAGAGAACTATC 572  
DB 181 AGGCTTGGATTGCTTTTGTATAGTGTCAAAATTTAGTGTGGAGAGAGAACTATC 240  
QY 573 AATGGCAATGAGACAGATGATGGGCAAGTCTTGGCAAAATTAATGACTGCCATGC 632  
DB 241 AATGGCAATGAGACAGATGATGGGCAAGTCTTGGCAAAATTAATGACTGCCATGC 300  
QY 633 AGGATGACCAACGCGCTTAACTTCTGCAATGCAAAAATTGAAAGTAATATCTA 692  
DB 301 AGGATGACCAACGCGCTTAACTTCTGCAATGCAAAAATTGAAAGTAATATCTA 360  
QY 693 AAAGTAAATATGACACAAATTCATATTAATTTGATGATGACTAATGTTGTCGT 752  
DB 361 AAAGTAAATATGACACAAATTCATATTAATTTGATGATGACTAATGTTGTCGT 420  
QY 753 TCAATTTGATGATCAATGCTTCCAGCAAGAGCCCAATCTGATGAGTCCATGATCA 812  
DB 421 TCAATTTGATGATCAATGCTTCCAGCAAGAGCCCAATCTGATGAGTCCATGATCA 480  
QY 813 AATAGTCAATATATTCATAATCTGATATCTATTTATGGAACAGGTGATGATTTTCA 872  
DB 481 AATAGTCAATATATTCATAATCTGATATCTATTTATGGAACAGGTGATGATTTTCA 540  
QY 873 ATGTTTCTGGATCTCAAAATGTCAGGCCCAAAATTTATCTTGTGTCCAGGTATGT 932  
DB 541 ATGTTTCTGGATCTCAAAATGTCAGGCCCAAAATTTATCTTGTGTCCAGGTATGT 600  
QY 933 ATAGTATGGAAGTATGAGATCTGCAAAATTCAGAACTATGATGCTATGTTACTGTA 992  
DB 601 ATAGTATGGAAGTATGAGATCTGCAAAATTCAGAACTATGATGCTATGTTACTGTA 660  
QY 993 AATGAAGCCCAAAATTTATCGGTGCCGAAATGAGATTAGATCAAGCTTGGCAGGAGGA 1052  
DB 661 AATGAAGCCCAAAATTTATCGGTGCCGAAATGAGATTAGATCAAGCTTGGCAGGAGGA 720  
QY 1053 TCTGGACAGCTTGCAACATCAAAATTTCTGATGATGGAATGCAAGACGTTAAATATCC 1112  
DB 721 TCTGGACAGCTTGCAACATCAAAATTTCTGATGATGGAATGCAAGACGTTAAATATCC 780  
QY 1113 ATATAT 1118  
DB 781 ATATAT 786

RESULT 3  
BM412528  
LOCUS  
DEFINITION  
EST86855 tomato breaker fruit Lycopersicon esculentum cDNA clone  
ACCESSION  
BM412528  
VERSION  
BM412528.1 GI:18264158  
KEYWORDS  
EST.  
SOURCE  
ORGANISM  
Lycopersicon esculentum (tomato)  
Bukharova; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 767)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karanicheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Uteback,T., Van Aken,S.,  
Roming,C.M., Fraser,C.M., Martin,G.B., Tanksey,S.D. and  
Giovannoni,J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.

FEATURES  
source  
Location/Qualifiers  
1..767  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA96"  
/db\_xref="taxon:4081"  
/clone="CIB6014"  
/tissue\_type="Pericarp"  
/dev\_stage="breaker"  
/lab\_pos="SOLR"  
/clone\_1lb="tomato breaker fruit"  
/note="Vector: pBluescriptSkcmudap; Site\_1: EcoRI;  
Site\_2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN  
Query Match 46.8%; Score 765.4; DB 12; Length 767;  
Best Local Similarity 99.9%; Pred. No. 3.7e-118;  
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 29 CCATACCATATTAACAATATATCATGCTTATCCAAAGAAATGATTTCTCTTCATTAAT 88  
DB 1 CCATACCATATTAACAATATATCATGCTTATCCAAAGAAATGATTTCTCTTCATTAAT 60  
QY 89 TATTTTGGTTCATCAATTTCAACTTGTAGAACAAATGTTATGATGACAAATTTATTCAA 148  
DB 61 TATTTTGGTTCATCAATTTCAACTTGTAGAACAAATGTTATGATGACAAATTTATTCAA 120  
QY 149 ACAAGTTTATGATTAATATCTTGAACAGAAATTTGCTCATGATTTCAAGCTTATCTTC 208  
DB 121 ACAAGTTTATGATTAATATCTTGAACAGAAATTTGCTCATGATTTCAAGCTTATCTTC 180  
QY 209 TATTTTGAACAAAATATTTGAAGCAACATATATTTGACAGGTTGATTAATATGAGAT 268  
DB 181 TATTTTGAACAAAATATTTGAAGCAACATATATTTGACAGGTTGATTAATATGAGAT 240  
QY 269 TAAAGTATTAATGATCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 328  
DB 241 TAAAGTATTAATGATCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 300  
QY 329 TGCATTTGAGCAAGCATGATGAGCAATGCTTATAGAACACCTGTTCAATTTGGT 388  
DB 301 TGCATTTGAGCAAGCATGATGAGCAATGCTTATAGAACACCTGTTCAATTTGGT 360  
QY 389 TCCATAAAAACAGAAATTAATCTTCTCAAGCAATCACTTTTCAGTTCATGAGATCTTC 448  
DB 361 TCCATAAAAACAGAAATTAATCTTCTCAAGCAATCACTTTTCAGTTCATGAGATCTTC 420  
QY 449 TATTTAGTAAAGATTTTGGATCTTATAGAGCATCTAGTAAATTTTGACATACAAAGA 508  
DB 421 TATTTAGTAAAGATTTTGGATCTTATAGAGCATCTAGTAAATTTTGACATACAAAGA 480  
QY 509 TGAAGGCTTGGATGCTTTTGAATAGTGTCAAAATTTAGTGTGAGAGAGAGAGAAC 568  
DB 481 TGAAGGCTTGGATGCTTTTGAATAGTGTCAAAATTTAGTGTGAGAGAGAGAGAAC 540  
QY 569 TATCATATGCAATGACAAATGATGATGAGCAAGTCTTCCAAATATTAATCACTGCC 628  
DB 541 TATCATATGCAATGACAAATGATGATGAGCAAGTCTTCCAAATATTAATCACTGCC 600  
QY 629 ATGAGGATGACCAACGCGCTTAACTTCTGCAATGCAAAAATTGAAAGTAATTA 688  
DB 601 ATGAGGATGACCAACGCGCTTAACTTCTGCAATGCAAAAATTGAAAGTAATTA 660  
QY 689 TCTAAAGATTAATAATGACAAACAAATTCATATCAATTTGAGTCAAGCACTAATGTTG 748  
DB 661 TCTAAAGATTAATAATGACAAACAAATTCATATCAATTTGAGTCAAGCACTAATGTTG 720  
QY 749 AGCTTCAATTTGATGATCATGCTTCCAGCAAGAGCCCAAAATATCTG 795



/lab host="SOLR"  
 /clone lib="tomato breaker fruit"  
 /note=Vector: pBluescriptSKmTadap1; Site 1: EcoR1;  
 Site 2: Xho1; supplier: Boyce Thompson Institute; fruit  
 sequencing: The Institute for Genomic Research. Fruit  
 were harvested at the breaker stage (first sign of  
 lycopersene accumulation on the blossom end of fruit). Fruits  
 were cut in half and the seeds and locules were discarded  
 prior to freezing the pericarp."

## ORIGIN

Query Match	45.4%	Score 742;	DB 12;	Length 742;
Best Local Similarity	100.0%	Pred. No. 3e-114;		
Matches 742;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	52	TGGTATCCAAAGGAATAGTATTCCTCCTTCATATTTATTTTGGCTCATCAATTCOA	11
Db	1	TGGTTATCCAAAGGAATAGTATTCCTCCTTCATATTTATTTTGGCTCATCAATTCOA	60
QY	112	CTTGTAAGCAATGTTATGTAGCAATTTATTCACAACAAGTTATGATTAATATTCCTG	171
Db	61	CTTGTAAGCAATGTTATGTAGCAATTTATTCACAACAAGTTATGATTAATATTCCTG	120
QY	172	AACAAGATTTGCTCATGATTTTTCAGGCTTACTCTTCTATTTGGACCAAAAATTTGAA	231
Db	121	AACAAGATTTGCTCATGATTTTTCAGGCTTACTCTTCTATTTGGACCAAAAATTTGAA	180
QY	232	GCAACAATAATATATGACAAGTTGATTAATAAATGGGATTAAGGATTAATGTACTAGCT	291
Db	181	GCAACAATAATATATGACAAGTTGATTAATAAATGGGATTAAGGATTAATGTACTAGCT	240
QY	292	TTGGAGCTAAGGGTGATGAAAAACAATATGATTAATATTCATTTGACAAGCATGGAATG	351
Db	241	TTGGAGCTAAGGGTGATGAAAAACAATATGATTAATATTCATTTGACAAGCATGGAATG	300
QY	352	AAGCATGTTCACTCTAGAACACCTGTTCAAATTTGTGTTCTTAAAAACAAGATTAATCTTC	411
Db	301	AAGCATGTTCACTCTAGAACACCTGTTCAAATTTGTGTTCTTAAAAACAAGATTAATCTTC	360
QY	412	TCAAGCAATCAACCTTTTCAGGTCATGAGACATCTTCTATTCAGTAAGATTTTGGAT	471
Db	361	TCAAGCAATCAACCTTTTCAGGTCATGAGACATCTTCTATTCAGTAAGATTTTGGAT	420
QY	472	CCTTAGAAGCATCTAGTAATAATTTTCAGACTACAAAGATTAAGAGGCTTTGGAATGCTTTG	531
Db	421	CCTTAGAAGCATCTAGTAATAATTTTCAGACTACAAAGATTAAGAGGCTTTGGAATGCTTTG	480
QY	532	ATAAGTTCOAAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGCAATGACAAAGTAT	591
Db	481	ATAAGTTCOAAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGCAATGACAAAGTAT	540
QY	592	GSTGGCCAAAGTCTTGCAAAATTAATTAATTCATCTGCATGCAAGGATGCAACACGGGCTT	651
Db	541	GSTGGCCAAAGTCTTGCAAAATTAATTAATTCATCTGCATGCAAGGATGCAACACGGGCTT	600
QY	652	TAACCTTTGGAATGCAAAAAATTTGAAGTGAATATCTTAAAGATTAATAAATGCAACAC	711
Db	601	TAACCTTTGGAATGCAAAAAATTTGAAGTGAATATCTTAAAGATTAATAAATGCAACAC	660
QY	712	AAATTCATATCAAAATTTGAGTCATGCACTAATGTTGTAGCTTCAAAATTTGATGATCAATG	771
Db	661	AAATTCATATCAAAATTTGAGTCATGCACTAATGTTGTAGCTTCAAAATTTGATGATCAATG	720
QY	772	CTTCAGCAAAAGGCCCAATATAC 793	
Db	721	CTTCAGCAAAAGGCCCAATATAC 742	

LOCUS	DEFINITION	FEATURES	ORIGIN	REFERENCE
BM413343	tomato breakear fruit Lycopersicon esculentum cDNA clone	749 bp mRNA linear EST 22-JAN-2007	CGEAG63C11 5' end, mRNA sequence.	

ACCESSION	EM413343
VERSION	EM413343.1
KEYWORDS	GI:18264973
SOURCE	EST.
ORGANISM	<i>Lycopersicon esculentum</i> (tomato)
REFERENCE	<i>Lycopersicon esculentum</i>
AUTHORS	Ekateriya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 749)
	Alcalá, J., Vrebalov, J., White, R., Václav, T., Karamycheva, S. A.,

/clone.lib="tomato breaker fruit"  
/note="vector: paluscriptskmtuadpt: Site 1: EcoRI;  
Site\_2: XhoI; supplier: Boyce Thompson Institute; Fruit  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN

Query March	45.1%	Score 738;	DB 112;	Length 749;
Best Local Similarity	99.9%	Pred. No. 1.4e-113;		
Matches 749;	Conservative	0;	Mismatches	0;
			Indels	1;
			Gaps	1;

Qy	268	TTAAAGAGATTAATGACTACTACCTTTGGAGCTAAAGGATGGAAAAATATGATATA	32
Dp	1	TTAAAGTATTANTGACTTACTCTTTGAGCTAAAGGATGGAAAAATATGATATA	60
Qy	328	TTGCATTTGAGCAAGCATGAAATGAAGCATGTTCATCTAGAACACCTGTTCAATTTG	387
Dp	61	TTGCATTTGAGCAAGCATGAAATGAAGCATGTTCATCTAGAACACCTGTTCAATTTG	120
Qy	388	TTTCTAAAAACAGATTTATCTTCACAGCAAATACCTTTGACAGTCACAGATCTT	447
Dp	121	TTTCTAAAAACAGATTTATCTTCACAGCAATATCCTTTGAGTCACAGATCTT	180
Qy	448	CTATTTCACTAAAGATTTTGGATCTTGAAGCATCTAGTAAATTTCAGACTACAAAG	507
Dp	181	CTATTTCACTAAAGATTTTGGATCTTGAAGCATCTAGTAAATTTCAGACTACAAAG	240
Qy	508	ATGAAAGGCTTTGGATGCTTTTGAATAGTGTCAAAATTTAGTGTGGAGAGAGGAA	567
Dp	241	ATGAAAGGCTTTGGATGCTTTTGAATAGTGTCAAAATTTAGTGTGGAGAGAGGAA	300
Qy	568	CTATCAATGGCATGACAAGATATGCTGGCCAAAGTCTTGCAAAATTAATTAATCATCTGC	627
Dp	301	CTATCAATGGCATGACAAGATATGCTGGCCAAAGTCTTGCAAAATTAATTAATCATCTGC	360
Qy	628	CATGCAAGGATGACCAACGAGCTTAACTTTCTGGAAATTCGAAAATTTGAAGTGAATA	687

Db	Accession	Source	Organism	Reference	Title	Journal	Comment
Db	361	CATCAGGAGTACGACCAACGGCCTTAACCTTCTGGAAATGGCAAAAATTGGAAGTAA	728 bp mRNA linear EST 22-JAN-2002	BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	688	ATCTAAAGAGTAAATAAGTCAACAACAAATTCATATCAATTTGAGCATGCACTAATGTG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	421	ATCTAAAGAGTAAATAAGTCAACAACAAATTCATATCAATTTGAGCATGCACTAATGTG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	748	TAGCTTCAAAATTTGATGATCAATGCTTCAGCAAAAGAGCCAAATCTATGAGATCCATG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	481	TAGCTTCAAAATTTGATGATCAATGCTTCAGCAAAAGAGCCAAATCTATGAGATCCATG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	808	TATCAAAATCTCAATATATTCGAAATATCTGATATTTATTGGAAACAGGTGATGTTGA		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	541	TATCAAAATCTCAATATATTCGAAATATCTGATATTTATTGGAAACAGGTGATGTTGA		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	868	TTTCAATTTGTTTCTGATCTCAAAAATGTCAGAGCCCAAAATATTTACTTGTGTCCAGTCT		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	601	TTTCAATTTGTTTCTGATCTCAAAAATGTCAGAGCCCAAAATATTTACTTGTGTCCAGTCT		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	928	ATGCTATATGATTTGAGAGCTTAGATATCGAAATTCAGAGCTATGTCTAATGTGA		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	661	ATGCTATATGATTTGAGAGCTTAGATATCGAAATTCAGAGCTATGTCTAATGTGA		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Qy	988	CTGTAAATGAAAGCCAAATTTATCGGTGCGG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
Db	720	CTGTAAATGAAAGCCAAATTTATCGGTGCGG		BM413075	Tomato breaker fruit Lycopersicon esculentum cDNA clone		
RESULT 7							
LOCUS	BM413075	728 bp	mRNA	linear	EST 22-JAN-2002		
DEFINITION	EST587402	tomato breaker fruit	Lycopersicon esculentum	cDNA clone			
ACCESSION	BM413075						
VERSION	BM413075.1	GI:18264705					
KEYWORDS							
SOURCE							
ORGANISM			Lycopersicon esculentum (tomato)				
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
COMMENT							
FEATURES							
Source							
Location/Qualifiers							
1..728							
/organism="Lycopersicon esculentum"							
/mol_type="mRNA"							
/cultivar="TA496"							
/db_xref="taxon:4081"							
/clone="cLBG62K21"							
/tissue_type="Pericarp"							
/dev_stage="Breaker"							
/lab_host="SOLR"							
/clone_id="tomato breaker fruit"							
/note="Vector: pBluescriptsmc2adpt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."							

ORIGIN	Query Match	44.4%	Score 726.4	DB 12	Length 728	
	Best Local Similarity	99.9%	Pred. No. 1.2e-111			
	Matches 727	Conservative 0	Mismatches 1	Indels 0	Gaps 0	
QY	787	CAAAATCTGATGGAGTCCATGATACAAATCTCAATATATTCAAATATCTGATACATCTTA	846			
Db	1	CAAAATCTGATGGAGTCCATGATACAAATCTCAATATATTCAAATATCTGATACATCTTA	60			
QY	847	TTGGACAGGTGATGATTTGATTTCAATTTGTTCTGGATCTCAAAATGTGACGGCCAA	906			
Db	61	TTGGACAGGTGATGATTTGATTTCAATTTGTTCTGGATCTCAAAATGTGACGGCCAA	120			
QY	907	ATATTTACTTTGTGTCCAGGTCTCATGTATTAAGTATTTGGAAGCTTAGATCTTGAAATTCAG	966			
Db	121	ATATTTACTTTGTGTCCAGGTCTCATGTATTAAGTATTTGGAAGCTTAGATCTTGAAATTCAG	180			
QY	967	AAGCTTATGTGTCTAATGTTACTGTAAATGAAGCCAAATATATGTGGCCGAAATGTGAG	1026			
Db	181	AAGCTTATGTGTCTAATGTTACTGTAAATGAAGCCAAATATATGTGGCCGAAATGTGAG	240			
QY	1027	TTAGGATCAAGACTTGGCAGGAGAGATCTGACAAAGCTAGCACATCAAAATTTCTGAATG	1086			
Db	241	TTAGGATCAAGACTTGGCAGGAGAGATCTGACAAAGCTAGCACATCAAAATTTCTGAATG	300			
QY	1087	TGGAAATGCAAGACGTTAAATATATCCCAATATTTATAGCCAAACTATTTGATTCGATG	1146			
Db	301	TGGAAATGCAAGACGTTAAATATATCCCAATATTTATAGCCAAACTATTTGATTCGATG	360			
QY	1147	AACCATTTATACAAAGCTTTTCAGCAGTTCAAGTGAAGAAATGTGGTATGAGAAATATCA	1206			
Db	361	AACCATTTATACAAAGCTTTTCAGCAGTTCAAGTGAAGAAATGTGGTATGAGAAATATCA	420			
QY	1207	AGGGCACAAGTGCACAAAGGTGGCCATAAATTTGATTTGACAGCACAACTTTTCATGTG	1266			
Db	421	AGGGCACAAGTGCACAAAGGTGGCCATAAATTTGATTTGACAGCACAACTTTTCATGTG	480			
QY	1267	AAGGAATTAATAATGGAGAAATATAATTTAGTAGGGGAAAGTGGAAACCATACAGAGCTA	1326			
Db	481	AAGGAATTAATAATGGAGAAATATAATTTAGTAGGGGAAAGTGGAAACCATACAGAGCTA	540			
QY	1327	CGTCCAAAATATGTCATTTTAAACATGCTGAACATGTTACACCACTGACACTTCACTAG	1386			
Db	541	CGTCCAAAATATGTCATTTTAAACATGCTGAACATGTTACACCACTGACACTTCACTAG	600			
QY	1387	AAATTTCCAGAGATGAGAGCTCTTTTGATATATTAATTAATTTATACATATAGACTTCCATA	1446			
Db	601	AAATTTCCAGAGATGAGAGCTCTTTTGATATATTAATTAATTTATACATATAGACTTCCATA	660			
QY	1447	TATAGCAGATATGATATATATCAACAATAAACAATCTATATCTATGATATTAATAATTA	1506			
Db	661	TATAGCAGATATGATATATATCAACAATAAACAATCTATATCTATGATATTAATAATTA	720			
QY	1507	TTAATATG 1514				
Db	721	TTAATATG 728				

RESULT 8  
AM222257  
LOCUS  
DEFINITION  
EST1293068 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA  
clone GLEN7018, mRNA sequence.  
AM222257  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

AM222257 729 bp mRNA linear EST 18-MAY-2001  
EST1293068 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA  
clone GLEN7018, mRNA sequence.  
AM222257  
AM222257.1 GI:6533941  
EST.  
Lycopersicon esculentum (tomato)  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 729)





```
Db 61 GATCAATCTTACGAGGAGCCCAATATCATGATGAGTCCATGATCAATATCATTA 120
Qy 824 TATTCAAATATCTGATATATATATATGGAACAGGTGATGATTTCAATTTGTTGCG 883
Db 121 TATTCAAATATCTGATATATATATATGGAACAGGTGATGATTTCAATTTGTTGCG 180
Qy 884 ATTCCAAAATGTGAGGAGCCCAAAATATATATCTTGTGTCAGGTGATGATTTGAG 943
Db 181 ATTCCAAAATGTGAGGAGCCCAAAATATATCTTGTGTCAGGTGATGATTTGAG 240
Qy 944 AAGCTTAGATCTGGAATTCAGAAAGTTATGTGTCTAATGTTACTGTAATGAGCCAA 1003
Db 241 AAGCTTAGATCTGGAATTCAGAAAGTTATGTGTCTAATGTTACTGTAATGAGCCAA 300
Qy 1004 AATATATGTCGTCGCAAAATGAGATGATGATGATGATGATGATGATGATGATGAT 1063
Db 301 AATATATGTCGTCGCAAAATGAGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 1064 TAGCAACATCAAAATTTCTGAATGTGAATGCAAGACCTTAAGTATCCATATATAGA 1123
Db 361 TAGCAACATCAAAATTTCTGAATGTGAATGCAAGACCTTAAGTATCCATATATAGA 420
Qy 1124 CCAAAATCTATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1183
Db 421 CCAAAATCTATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 1184 AAATGTGTCGTCGCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1243
Db 481 AAATGTGTCGTCGCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Qy 1244 TTGAGAGCAAAATCTTCCATGTGAAGAAATATATATGAGAAATATATATATAGA 1303
Db 541 TTGAGAGCAAAATCTTCCATGTGAAGAAATATATATGAGAAATATATATATAGA 600
Qy 1304 AAGTGAAGAAATCTATGAGGCTAGCTGCAAAATGTCATTTTAAATGCTGAACATGT 1363
Db 601 AAGTGAAGAAATCTATGAGGCTAGCTGCAAAATGTCATTTTAAATGCTGAACATGT 660
Qy 1364 TACACCACTGACCTGACCTAGTAAATTTGAGAGATGAGTCAAGCTCTTTTG 1412
Db 661 TACACCACTGACCTGACCTAGTAAATTTGAGAGATGAGTCAAGCTCTTTTG 709

RESULT 10
AM442253 696 bp mRNA linear EST 18-MAY-2001
LOCUS EST311649 tomato fruit red ripe, TMU Lycopersicon esculentum cDNA
DEFINITION
AM442253
VERSION AM442253
KEYWORDS
SOURCE
ORGANISM Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Eubacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 696)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upson,J., Hansen,T., Craven,M.B., Bowman,C.L., Alm,S.,
Romling,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
REFERENCE Generation of ESTs from tomato fruit tissue
Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..696
/organism="Lycopersicon esculentum"
```

```
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEN22614"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/clone_lib="tomato fruit red ripe, TMU"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopen accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

ORIGIN
Query Match 42.4%; Score 694.4; DB 10; Length 696;
Best Local Similarity 99.9%; Pred. No. 2.7e-106;
Matches 695; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 349 ATGAAGCATGTTTCATCTAGAAACCTGTTCAATTTGTTGTTCTTAAACAGAAATATTC 408
Db 1 ATGAAGCATGTTTCATCTAGAAACCTGTTCAATTTGTTGTTCTTAAACAGAAATATTC 60
Qy 409 TTCTCAAGCAATCACTTTTCAGGTCATGACAGTCTTCTATTTCAATTAAGATTTTG 468
Db 61 TTCTCAAGCAATCACTTTTCAGGTCATGACAGTCTTCTATTTCAATTAAGATTTTG 120
Qy 469 GATCCTTAGAAGCATCTAGTAAATTTTGAAGTGAATGAAGGCTTTGATTTGCTT 528
Db 121 GATCCTTAGAAGCATCTAGTAAATTTTGAAGTGAATGAAGGCTTTGATTTGCTT 180
Qy 529 TTGATAGTTTCAAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGAATGAGCAAG 588
Db 181 TTGATAGTTTCAAAATTTAGTTGTTGAGAGAGAGAACTATCAATGGAATGAGCAAG 240
Qy 589 TATGTTGGCCAAAGTTCTTGCAAAATTAATTAATCACTCCATGAGGAGATGACCAAGG 648
Db 241 TATGTTGGCCAAAGTTCTTGCAAAATTAATTAATCACTCCATGAGGAGATGACCAAGG 300
Qy 649 CCTTAACTTCTGGAATTCGAAAAATTTGAAGTGAATATCTAAAGATGAATGATGAC 708
Db 301 CCTTAACTTCTGGAATTCGAAAAATTTGAAGTGAATATCTAAAGATGAATGATGAC 360
Qy 709 AACCAATTCATATCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 768
Db 361 AACCAATTCATATCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 769 ATGCTTCAAGCAAGGAGCCCAATCTGATGAGTGCATGATCAAAATCTCAATATATTC 828
Db 421 ATGCTTCAAGCAAGGAGCCCAATCTGATGAGTGCATGATCAAAATCTCAATATATTC 480
Qy 829 AAATATCTGATATATATATGAAACAGGTGATGATGATGATGATGATGATGATGATGAT 888
Db 481 AAATATCTGATATATATATGAAACAGGTGATGATGATGATGATGATGATGATGATGAT 540
Qy 889 AAAATGTGAGGAGCCCAAAATTTCTTGTGTCAGGTGATGATGATGATGATGATGATGAT 948
Db 541 AAAATGTGAGGAGCCCAAAATTTCTTGTGTCAGGTGATGATGATGATGATGATGATGAT 600
Qy 949 TAGATCTGGAATTCGAAAGCTTATGTCATGATGATGATGATGATGATGATGATGATGAT 1008
Db 601 TAGATCTGGAATTCGAAAGCTTATGTCATGATGATGATGATGATGATGATGATGATGAT 660
Qy 1009 TCGTGGCCGAAATGAGGTTAGATCAAGACTTGGC 1044
Db 661 TCGTGGCCGAAATGAGGTTAGATCAAGACTTGGC 696

RESULT 11
AM442335
```

LOCUS AM442335 687 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST111731 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA  
ACCESSION AM442335  
VERSION AM442335  
KEYWORDS EST.  
SOURCE GI:6977586  
ORGANISM Lycopersicon esculentum (tomato)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 687)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,  
Romling,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and  
Giovannoni,J.  
TITLE Generation of ESTs from tomato fruit tissue  
JOURNAL Unpublished (1999)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..687  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLN22113"  
/tissue\_type="pericarp"  
/dev\_stage="red ripe (7-20 days post-breaker)"  
/clone\_lib="tomato fruit red ripe, TAMU"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Giovannoni; Fruit were tagged at the  
breaker stage (first sign of lycopen accumulation on the  
blossom end of the fruit) and harvested 7 days  
post-breaker (fully red-ripe), 10 days post breaker, and  
20 days post-breaker (over-ripe). 20 day fruit which  
showed external or internal signs of pathogenesis were  
discarded. Fruit were cut in half and the seeds and  
locules were discarded prior to freezing the pericarp."

ORIGIN  
Query Match 41.8%; Score 683.8; DB 10; Length 687;  
Best Local Similarity 99.7%; Pred. No. 1.6e-104;  
Matches 685; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 274 TGATTATGACTTAGCTTTGAGCTAAGGATGAGAAACATATGATATATTCAT 333  
DB 1 TGATTATGACTTAGCTTTGAGCTAAGGATGAGAAACATATGATATATTCAT 60  
QY 334 TTGAGCAAGCATGGAATGAAGCATGTCATAGAACACCTGTCATTTGGTCTTA 393  
DB 61 TTGAGCAAGCATGGAATGAAGCATGTCATAGAACACCTGTCATTTGGTCTTA 120  
QY 394 AAAACAAGATTTCTTCTCAAGCAATACCTTTGAGTCATGAGATCTTCAATT 453  
DB 121 AAAACAAGATTTCTTCTCAAGCAATACCTTTGAGTCATGAGATCTTCAATT 180  
QY 454 CAGTAAGAATTTTGGATCCTTAGAAGCATCTAGTAATAATTTCAGACTCAAGATAGAA 513  
DB 181 CAGTAAGAATTTTGGATCCTTAGAAGCATCTAGTAATAATTTCAGACTCAAGATAGAA 240  
QY 514 GCGTTTGATGCTTTTGGATGCTTCAAAATTTAGTTGGAGAGAGAACTATCA 573  
DB 241 GCGTTTGATGCTTTTGGATGCTTCAAAATTTAGTTGGAGAGAGAACTATCA 300  
QY 574 ATGGCAATGGAAGAATAGTGGCAAGTCTTGCAAAATTAATTAATCACTGCGATGCA 633  
DB 301 ATGGCAATGGAAGAATAGTGGCAAGTCTTGCAAAATTAATTAATCACTGCGATGCA 360

QY 634 GGGATGCACCAACGGCCTTAACCTTGTGCAATTTGCAGAAATTTGAAAGTGAATATCTAA 693  
DB 361 GGGATGCACCAACGGCCTTAACCTTGTGCAATTTGCAGAAATTTGAAAGTGAATATCTAA 420  
QY 694 AGAGTAAATATGACACAAATTCATATCAATTTAGTCATGACCTAATGTGTGCTT 753  
DB 421 AGAGTAAATATGACACAAATTCATATCAATTTAGTCATGACCTAATGTGTGCTT 480  
QY 754 CAATTTGATGATCAATGCTTCAGCAAGAGCCCAATTAATCTGATGAGTCCATATCA 813  
DB 481 CAATTTGATGATCAATGCTTCAGCAAGAGCCCAATTAATCTGATGAGTCCATATCA 540  
QY 814 ATACTCAATATATTCATATATCTGATATCTATTTGAGAACAGGTGATGATTTGAA 873  
DB 541 ATACTCAATATATTCATATATCTGATATCTATTTGAGAACAGGTGATGATTTGAA 600  
QY 874 TTGTTCTGATCTCAAAATGTCAGAGCCCAAAATTTCTGTGTCCAGGTCAATGTA 933  
DB 601 TTGTTCTGATCTCAAAATGTCAGAGCCCAAAATTTCTGTGTCCAGGTCAATGTA 660  
QY 934 TAAGTATGGAAGCTTAGATCTGAA 960  
DB 661 TAAGTATGGAAGCTTAGATCTGAA 687

RESULT 12  
BM536089 703 bp mRNA linear EST 20-FEB-2002  
LOCUS BM536089  
DEFINITION EST589111 tomato breaker fruit Lycopersicon esculentum cDNA clone  
CLB671174 5' end, mRNA sequence.  
ACCESSION BM536089  
VERSION BM536089.1 GI:18615229  
KEYWORDS EST.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 703)  
Alcala,J., Vrebalov,J., White,R., Viston,T., Karameyheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Uteyback,T., Van Aken,S.,  
Romling,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and  
Giovannoni,J.  
TITLE Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
JOURNAL Unpublished (2002)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3  
FEATURES  
source  
1..703  
Location/Qualifiers  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLB671174"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSkudnap; Site\_1: EcoRI;  
Site\_2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopen accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN



Db 601 AGGCAACAAGAGTGGCCATTAATTGATTCAGCACAACACTTCCTCATGTGAAGAAATT 660  
QY 1275 ATATAGAGAAATATATA 1291  
Db 661 ATATAGAGAAATATATA 677

RESULT 14  
BM408546 675 bp mRNA linear EST 22-JAN-2002  
LOCUS EST582893 tomato breaker fruit Lycopersicon esculentum cDNA clone  
DEFINITION CLEB57K1 5' end, mRNA sequence.  
ACCESSION BM408546 GI:18260176  
VERSION EST.  
KEYWORDS Lycopersicon esculentum (tomato)  
SOURCE Lycopersicon esculentum  
ORGANISM Lycopersicon esculentum (tomato)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 675)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Utecherback,T., Van Aken,S.,  
Romano,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and  
Giovannoni,J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3

FEATURES  
source  
Location/Qualifiers  
1..675  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEB57K1"  
/tissue\_type="Pericarp"  
/dev\_stage="Breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSkmTadap; Site 1: EcoRI;  
Site 2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN  
Query Match 41.0%; Score 670.2; DB 12; Length 675;  
Best Local Similarity 99.6%; Pred. No. 3e-102;  
Matches 672; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 359 TTATCTAGAACACCTGTTCAATTTGTGTTCTTAATAAACAAGATTATCTTCAAGCA 418  
Db 1 TTATCTAGAACACCTGTTCAATTTGTGTTCTTAATAAACAAGATTATCTTCAAGCA 60  
QY 419 AATCACCCTTTGAGTCATGACAGATCTTATTTGCTGAAGATTTTGGATCCTTAGA 478  
Db 61 AATCACCCTTTGAGTCATGACAGATCTTATTTGCTGAAGATTTTGGATCCTTAGA 120  
QY 479 AGCATCTAGTAAATTTCAAGACTACAAAGATAGAGGCTTTGATGCTTTGATAGTGT 538  
Db 121 AGCATCTAGTAAATTTCAAGACTACAAAGATAGAGGCTTTGATGCTTTGATAGTGT 180  
QY 539 TCAAAATTGATGTTGGAGAGAGAACTATTCATGCAATGAGCAAGTATGTTGGCC 558

Db 181 TCAAAATTGATGTTGGAGAGAGAACTATTCATGCAATGAGCAAGTATGTTGGCC 240  
QY 599 AAGTCTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 658  
Db 241 AAGTCTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 300  
QY 659 CTGGAATTGCAAAATTTGAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 718  
Db 301 CTGGAATTGCAAAATTTGAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 360  
QY 719 TATCAAAATTGAGTCAAGCACTAATGTTAGCTTCAAAATTTGATGATCAATGCTTCAAGC 778  
Db 361 TATCAAAATTGAGTCAAGCACTAATGTTAGCTTCAAAATTTGATGATCAATGCTTCAAGC 420  
QY 779 AAAGAGCCCAATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 838  
Db 421 AAAGAGCCCAATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 480  
QY 839 TACTATTATTGGAACAGGTGATGATTGATTTCAATTTGTTCTGATCTCAAAATGTGCA 898  
Db 481 TACTATTATTGGAACAGGTGATGATTGATTTCAATTTGTTCTGATCTCAAAATGTGCA 540  
QY 899 GGCCACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 958  
Db 541 GGCCACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600  
QY 959 AATTCAGAACTTATGTTCTAATGTTAATGTTAATGAAAGCCAAATTAATGATGCGTCCGA 1018  
Db 601 AATTCAGAACTTATGTTCTAATGTTAATGTTAATGAAAGCCAAATTAATGATGCGTCCCG 660  
QY 1019 AATTCAGAACTTATGTTCTAATGTTAATGTTAATGAAAGCCAAATTAATGATGCGTCCCG 1033  
Db 661 AATTCAGAACTTATGTTCTAATGTTAATGTTAATGAAAGCCAAATTAATGATGCGTCCCG 675

RESULT 15  
BM411547 670 bp mRNA linear EST 22-JAN-2002  
LOCUS EST585874 tomato breaker fruit Lycopersicon esculentum cDNA clone  
DEFINITION CLEB57K1 5' end, mRNA sequence.  
ACCESSION BM411547  
VERSION BM411547.1 GI:18263177  
KEYWORDS EST.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 670)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Utecherback,T., Van Aken,S.,  
Romano,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and  
Giovannoni,J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.

FEATURES  
source  
Location/Qualifiers  
1..670  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEB57K1"  
/tissue\_type="Pericarp"  
/dev\_stage="Breaker"  
/lab\_host="SOLR"



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2004, 17:04:52 ; Search time 28.379 Seconds  
(without alignments)  
831.356 Million cell updates/sec

Title: US-10-018-604-2

Perfect score: 2390  
Sequence: 1 MWIQRNSILLIIIFASIS.....VPHCTSLISEDEALYNY 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*
  - 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*
  - 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*
  - 5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*
  - 6: /cgn2\_6/prodata/2/iaa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2390	100.0	457	6	5447867-1 Patent No. 5447867
2	1113	46.6	433	4	US-09-051-239A-2 Sequence 2, App11
3	1103	46.2	433	3	US-08-941-532-6 Sequence 6, App11
4	787	32.9	514	3	US-08-467-023-134 Sequence 134, App Patent No. 5453566
5	365	15.3	70	6	5453566-2 Sequence 189, App
6	319	13.3	127	3	US-08-467-023-189 Sequence 6652, App
7	299	12.5	451	4	US-09-107-532A-6652 Sequence 5, App11
8	298.5	12.5	452	1	US-08-290-878A-5 Sequence 2, App11
9	298.5	12.5	452	2	US-08-780-869-5 Sequence 5, App11
10	293.5	12.3	362	4	US-09-542-767A-1 Sequence 6, App11
11	257.5	10.8	360	3	US-08-061-062A-6 Sequence 1, App11
12	241	10.1	440	1	US-08-061-062A-8 Sequence 8, App11
13	241	10.1	440	1	US-08-061-062A-8 Sequence 8, App11
14	241	10.1	440	3	US-08-536-150-6 Sequence 6, App11
15	241	10.1	440	3	US-08-536-150-8 Sequence 6, App11
16	226.5	9.5	127	3	US-08-467-023-188 Sequence 188, App
17	212.5	8.9	442	4	US-09-107-532A-6254 Sequence 6254, App
18	205.5	8.6	415	4	US-09-198-956-6 Sequence 6, App11
19	205.5	8.6	415	4	US-09-670-141-6 Sequence 187, App
20	188.5	7.9	128	3	US-08-467-023-187 Sequence 8, App11
21	187	7.8	49	3	US-08-941-532-8 Sequence 41, App11
22	129.5	5.4	1005	4	US-09-206-942-41 Sequence 8, App11
23	129.5	5.4	1011	4	US-09-206-942-39 Sequence 39, App11
24	124.5	5.2	901	4	US-09-134-001C-5351 Sequence 5351, App
25	120.5	5.0	1004	4	US-09-206-942-57 Sequence 57, App11
26	120.5	5.0	1010	4	US-09-206-942-55 Sequence 55, App11
27	120	5.0	1535	3	US-08-755-587-185 Sequence 185, App

28	118.5	5.0	1073	4	US-09-206-942-49	Sequence 49, App1
29	118.5	5.0	1079	4	US-09-206-942-47	Sequence 47, App1
30	117	4.9	378	4	US-09-134-001C-4692	Sequence 4692, App
31	116.5	4.9	977	4	US-09-206-942-53	Sequence 53, App1
32	116.5	4.9	983	4	US-09-206-942-51	Sequence 51, App1
33	115	4.8	942	4	US-09-489-039A-13982	Sequence 13982, App
34	114	4.8	1036	4	US-09-206-942-73	Sequence 73, App1
35	114	4.8	1477	4	US-09-206-942-71	Sequence 71, App1
36	113.5	4.7	969	4	US-09-206-942-32	Sequence 32, App1
37	113.5	4.7	975	4	US-09-206-942-30	Sequence 30, App1
38	112	4.7	1612	1	US-08-169-927-2	Sequence 2, App11
39	110	4.6	975	4	US-09-328-352-4764	Sequence 4764, App
40	110	4.6	1220	4	US-09-206-942-28	Sequence 28, App1
41	110	4.6	1226	4	US-09-206-942-26	Sequence 26, App1
42	109.5	4.6	1338	2	US-08-728-470-9	Sequence 9, App11
43	109.5	4.6	1338	3	US-08-719-641-9	Sequence 9, App11
44	109.5	4.6	1529	2	US-08-728-470-10	Sequence 10, App1
45	109.5	4.6	1529	3	US-08-719-641-10	Sequence 10, App1

ALIGNMENTS

RESULT 1									
5447867-1									
Patent No. 5447867									
APPLICANT: BRIDGES, IAN; SCHUCH, WOLFGANG; GRIERSON, DONALD									
TIME OF INVENTION: RECOMBINANT DNA CONTAINING PECTIN									
RESTRICTION GENE SEGMENTS									
NUMBER OF SEQUENCES: 4									
CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US/08/24,866									
FILING DATE: 26-FEB-1993									
PRIORITY APPLICATION DATA:									
APPLICATION NUMBER: 720,629									
FILING DATE: 25-JUN-1991									
APPLICATION NUMBER: 419,779									
FILING DATE: 29-SEP-1989									
APPLICATION NUMBER: 119,614									
FILING DATE: 12-NOV-1987									
SEQ ID NO:1:									
LENGTH: 457									
5447867-1									
Query Match 100.0%; Score 2390; DB 6; Length 457;									
Best Local Similarity 100.0%; Pred. No. 5.8e+208;									
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MW	IQRNSILLIIIFASIS	STCRSNVIDNLFQVYDNIIEFPADHPQALSYLSKNIE	60				
DB	1	MW	IQRNSILLIIIFASIS	STCRSNVIDNLFQVYDNIIEFPADHPQALSYLSKNIE	60				
QY	61	NN	NDIKVKNKGKIVNVIS	SPGAKGDKTYDNIAPQAWAEACSSRPQVFPKKNYL	120				
DB	61	NN	NDIKVKNKGKIVNVIS	SPGAKGDKTYDNIAPQAWAEACSSRPQVFPKKNYL	120				
QY	121	LK	ITSGPQRSSISYKIFG	SLFASISKISPYKDRIMIAFDSYQNLVWGGGTTINGQV	180				
DB	121	LK	ITSGPQRSSISYKIFG	SLFASISKISPYKDRIMIAFDSYQNLVWGGGTTINGQV	180				
QY	181	WW	SSCKINKSLPCRDPA	LTPALTPWNCNKLKVNMLKSKNAQOIHKEKSCNVAASNMIN	240				
DB	181	WW	SSCKINKSLPCRDPA	LTPALTPWNCNKLKVNMLKSKNAQOIHKEKSCNVAASNMIN	240				
QY	241	AS	AKSNPTGVHNSVNOY	IOISDTITGTGDDCISYSGSNVQATNTGPRGHISGSL	300				
DB	241	AS	AKSNPTGVHNSVNOY	IOISDTITGTGDDCISYSGSNVQATNTGPRGHISGSL	300				
QY	301	GG	NSNAYVSNVTVAEAK	ITGAENGVRKKTQGGSGQASNIKFLNEMQDVKPIIIDQN	360				
DB	301	GG	NSNAYVSNVTVAEAK	ITGAENGVRKKTQGGSGQASNIKFLNEMQDVKPIIIDQN	360				
QY	361	YC	DRVEPCCIQQGSAV	QVKNVVENIKGTSATKVAIKFDDCSITPFCGIIIMENINLVGSEG	420				
DB	361	YC	DRVEPCCIQQGSAV	QVKNVVENIKGTSATKVAIKFDDCSITPFCGIIIMENINLVGSEG	420				

Db 361 YCDREVEPCIOQFSAYOVNAYVENIKGTSATKVAIKFDCINFPCEGIMENINLVGSG 420  
QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEISEDBALLVNY 457  
Db 421 KPSEATCKNVHFNNAEHVTPHCTSLSEISEDBALLVNY 457

## RESULT 2

US-09-051-239A-2  
Sequence 2, Application US/09051239A  
Patent No. 6420628  
GENERAL INFORMATION:  
APPLICANT: ULVSKOV, Peter  
APPLICANT: CHILD, Robin  
APPLICANT: VAN ONCKELIN, Henri  
APPLICANT: PRINSEN, Els  
APPLICANT: BORKHARDT, Bernard  
APPLICANT: SANDER, Ialil  
APPLICANT: PETERSEN, Morten  
APPLICANT: BUNDGARD POULSEN, Gert  
APPLICANT: BOTTERMAN, Johan  
TITLE OF INVENTION: Seed Shattering  
FILE REFERENCE: 2121-0138P  
CURRENT APPLICATION NUMBER: US/09/051,239A  
CURRENT FILING DATE: 1998-09-28  
PCT/EP96/04313  
PRIOR FILING DATE: 1996-10-04  
PRIOR APPLICATION NUMBER: EP 95 402241.4  
PRIOR FILING DATE: 1995-10-06  
PRIOR APPLICATION NUMBER: EP 95 203328.0  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 433  
TYPE: PRT  
ORGANISM: Braessica napus  
FEATURE:  
OTHER INFORMATION: Strain cv. Topaz.  
US-09-051-239A-2

Query Match 46.6%; Score 1113; DB 4; Length 433;  
Best Local Similarity 49.9%; Pred. No. 2.2e-92;  
Matches 223; Conservative 69; Mismatches 121; Indels 34; Gaps 8;

QY 7 SILLIIIFASISITCRSNVIDNLFKQVYNNILBOEFAHFOAYLSLKNISNNND 66  
Db 9 AAVLCVILMLACCOALSSNV-DDG-----YGHEDG---SFESDSLTKLNDD 51  
QY 67 KV-----DKNGIKVINLVSPFGAKGDKTYDNIAPFOAMNEACSSRTPVQFVVPKKN 118  
Db 52 DVLLTKSDRPTTSSSTYSVSNFPAKDGKTDTPQAFKAKKACSTNGVTFILPKGKT 111  
QY 119 YLLKQITFSGPCRSSISVKTIFGSLFASISKISDY-KDRRLWIAPDSVONLVVGG--GTIN 175  
Db 112 YLLKSIIFRPGCKSLRSPFOILGTLSASTKRSYDNDKHWLILBVDVNNLSIDGSAQIVD 171  
QY 176 GNGQVWPPSSCKINKSLPCRPAPALTFMNCNKLKVNLSKNAQOQHIFESCTNVVAS 235  
Db 172 GNGNIMWQNSCKIDKSPCTAPALTLNKLNVKRLVRNAQOQIOLISTEKNNVGVK 231  
QY 236 NLMNINAKSPNTDGVHVSNTQYIISPTIIGTDDCISIVSGSONVQATNITGPGHGI 295  
Db 232 NVKLTAGDSNTGTGIRIVATKIRISNSDIGTDDCISIEDGSONVQINDITGPGHGI 291  
QY 296 SIGSLSGNSEAVYNSVTNVEAKIIGAENGVRITWOGSGQASNIKFLNVEMODVKYPI 355  
Db 292 SIGSLGDNASKAYYSIDVDGATLSETDNGVRITKYOGSGTAKNIKFNIRMDNVKNPI 351  
QY 356 IIDQNYCDREVEPCIOQFSAYOVNAYVENIKGTSATKVAIKFDCSTNFPCEGIMENINL 415  
Db 352 IIDQNYCDK-DKCBQOESAVOVNAYVQNIKGTISATDVALMNGSVKXPCGIVLENNVI 410

QY 416 VGESGKPESEATCKNVHFNNAEHVTPHC 442  
Db 411 KG-----GKASCENNVKDKGTVPK 432

## RESULT 3

US-08-941-532-6  
Sequence 6, Application US/08941532  
Patent No. 6096946  
GENERAL INFORMATION:  
APPLICANT: ROBERTS, Jeremy Alan  
APPLICANT: COUPE, Simon Allan  
APPLICANT: JENKINS, Elizabeth Sarah  
TITLE OF INVENTION: CONTROL OF POD DEHISCENCE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,532  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/00757  
FILING DATE: 29-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9506684.1  
FILING DATE: 31-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bsmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0623.0580001/RWE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 433 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-941-532-6

Query Match 46.2%; Score 1103; DB 3; Length 433;  
Best Local Similarity 49.8%; Pred. No. 1.7e-91;  
Matches 221; Conservative 70; Mismatches 125; Indels 28; Gaps 7;

QY 7 SILLIIIFASISITCRSNVID-----DNLFKQVYNNILBOEFAHFOAYLSYLSKNIES 61  
Db 9 AAVLCVILMLACCOALSSNVDDYGHEDGSFET--DSLTKLNDDVDVLLKSSDRPTTES 66  
QY 62 NNNIDYDKNGIKIKINLVSPFGAKGDKTYDNIAPFOAMNEACSSRTPVQFVVPKKNYL 121  
Db 67 S-----TVSVSNFPAKDGKTDTPQAFKAKKACSTNGVTFILPKGTYLL 114  
QY 122 KQITFSGPCRSSISVKTIFGSLFASISKISDY-KDRRLWIAPDSVONLVVGG--GTINNG 178  
Db 115 KSIFRPGCKSLRSPFOILGTLSASTKRSYDNDKHWLILBVDVNNLSIDGSAQIVDNG 174  
QY 179 QVWPPSSCKINKSLPCRPAPALTFMNCNKLKVNLSKNAQOQHIFESCTNVVASVLM 238  
Db 175 KIMWQNSCKIDKSPCTAPALTLNKLNVKRLVRNAQOQIOLISTEKNVSDVKNVK 234



```

QY 239 |NAKASKEPTDGVHVSNOYQIOJDTITCTGDDCISYSSGSONQANITNCTGPHGISG| 298
Db 235 |TAPDESDNTDGIHIVATKNIRISDITGDDCISIEDGSQNVQINDLTCTGPHGISG| 294
QY 299 |SLGSGNSAVYNSVNVNEAKIIGAEVGRITKMOGSGQAINIFELNVEMQDVXYPIIID| 358
Db 295 |SLGDDNSGAVYNSGVINDATITSEDDNGVRLITYQSGSTACKNITFQNRMDNVZNPRIID| 354
QY 359 |QNYCDRVEPTIQPSAVQVKNVYENIKGTSATKVAIKFDCTNPFEGILMENINIVGE| 418
Db 355 |QNYCDK-DKCGQGSQSAVQVNVNVYENIKGTSATVAINFNSVXYPQGVILEVNIKIG-| 412
QY 419 |SGKPSKATCKVNHENNAEHVTPHC| 442
Db 413 |---GKASCKKNVNVMDKGTAVSPKC| 432

RESULT 4
US-08-467-023-134
/ Sequence 134, Application US/08467023
/ Patent No. 6090386
/ GENERAL INFORMATION:
/ APPLICANT: Griffith, Irwin J.;
/ APPLICANT: Pollock, Joanne;
/ APPLICANT: Bond, Julian F.;
/ APPLICANT: Garman, Richard D.;
/ APPLICANT: Kuo, Mei-Chang;
/ APPLICANT: Yeung, Siu-mei H.;
/ APPLICANT: Brauer, Andrew;
/ APPLICANT: Kexley, Mark A.;
/ APPLICANT: Powers, Steven P.
/ TITLE OF INVENTION: Allergenic Proteins And Peptides From
/ TITLE OF INVENTION: Japanese Cedar Pollen
/ NUMBER OF SEQUENCES: 261
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Immunologic Pharmaceutical Corporation, Inc.
/ STREET: 610 Lincoln St
/ CITY: Waltham
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02154

COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/467,023
/ FILING DATE: June 6, 1995
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/350,225
/ FILING DATE: December 6, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jane E. Remillard
/ REGISTRATION NUMBER: 38,872
/ REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 227-7400
/ TELEFAX: (617) 227-5941
/ INFORMATION FOR SEQ ID NO: 134:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 514 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-467-023-134

Query Match 32.9%; Score 787; DB 3; Length 514;
Best Local Similarity 40.5%; Pred. No. 9,1e-63;
Matches 168; Conservative 73; Mismatches 160; Indels 14; Gaps 7;
45 AHDFAYL---SYLSKNTESNNNDIKVD---KNGIKVIVNLSPFGAKGDKGKTYDNIATFEQA 98

```

[illegible]

APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Waltham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMT-028CPD2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 127 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-189

Query Match 13.3%; Score 319; DB 3; Length 127;  
Best Local Similarity 56.0%; Pred. No. 2,7e-21;  
Matches 65; Conservative 16; Mismatches 35; Indels 0; Gaps 0;

QY 293 HGISIGLSGNSPAAVSNVTNEAKIIGAENGVRITKMGSGQASNIKFLNTEMQDVX 352  
DB 8 HEFISIGLSGNSPAAVSNVTNEAKIIGAENGVRITKMGSGQASNIKFLNTEMQDVX 67

QY 353 YPIIIDNVCGRVPCIQGFSAVQKVVNENIGTSATKAKFDDSTNFPCEGI 408  
DB 68 NPIIINQFYCTASACQNRSAVQIDVTYKNIGTSATKAKFDDSTNFPCEGI 123

RESULT 7  
US-09-107-532A-6652  
Sequence 6652, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007  
INFORMATION FOR SEQ ID NO: 6652:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 451 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...451  
SEQUENCE DESCRIPTION: SEQ ID NO: 6652:  
US-09-107-532A-6652

Query Match 12.5%; Score 299; DB 4; Length 451;  
Best Local Similarity 27.2%; Pred. No. 1.1e-18;  
Matches 99; Conservative 65; Mismatches 132; Indels 68; Gaps 15;

QY 73 IKVINIVSFGAKGDKTYNDIAFEQAMNEACSSRTPVQFVKKKNYLLKQITFSGPCRS 132  
DB 5 IMWYDILTFGASID--ELNTEALIQDAIDAAASDGGGT-VVFPAGB-----FLTGALFL 54

QY 133 SISVKIFGSLPSSAKIS-DYKD-----RLW---IARDVQNLVVGSGGTG 176  
DB 55 KSNVELHLSGAVLAFKFDDEPDYVHVRMEGVHAKYVASCTIYQWVENISVTGGLTD 114

QY 177 NGQVWV-----PSCCKINKSLPCRDAPFALTFWNCKNLKYNLKSRAQQLHKFESCT 230  
DB 115 NGKKWHTFFRNEPD---NLAYP---RPKLMSPFNCHRTIVKDIKLIQSPMTINPIICS 167

QY 231 NVVASNLINNASAPPTDGVHVSNTQYIOISDTIIGTDDCISIVSGS-----QN 281  
DB 168 NATPDNLILNPADSPNDGIDPESCKNVRISNCHIDVGDCLAKGTEETTERIACEN 227

QY 282 VQATNITCGPH--GISISLGSNGSEAVSNVTNEAKIIGAENGVRITKMGSGQASN 340  
DB 228 ITITNCTMVHGHGVVLSGSEWGS---INRITISNCFQFTDGRILKSRGGIYED 283

QY 341 IKPLNEMQDVKYIITIDQNY-CD--RVEPCIQ-----PSAVQKVVYNI 385  
DB 284 IRVSNIVDMVCMFILLVLYFCGRGPEYVWKKAYPIDERTPAPRIHFNSITARNV 343

QY 386 KGTG 389  
DB 344 HASA 347

RESULT 8  
US-08-290-978A-5  
Sequence 5, Application US/08290978A  
Patent No. 5624834  
GENERAL INFORMATION:  
APPLICANT: KUSTERS-VAN SOMEREN, MARGO A.  
APPLICANT: MULLER, YVONNE

APPLICANT: KESTER, HERMANUS C.M.  
APPLICANT: VISSER, JACOB  
APPLICANT: VAN COYEN, ALBERT J.J.  
APPLICANT: ROLIN, CLAUD  
TITLE OF INVENTION: CLONING AND EXPRESSION OF THE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,978A  
FILING DATE: 17-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 4615-0044.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 452 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-290-978A-5

Query Match 12.5%; Score 298.5; DB 1; Length 452;  
Best Local Similarity 25.1%; Pred. No. 1.3e-18;  
Matches 107; Conservative 73; Mismatches 166; Indels 81; Gaps 16;

74 KVINLVSFGAKDGKTYDNIAPFOANNEACSSRTPVQFVVPKKNYLL--KQITFGSPC 130  
48 KTCVASHGEGTDDSYIISALNQ-----CNHGKVF--DEKKEYIIGALNMTF---- 96  
131 RSSISVKIFGSL-----EASSKISDYKDRRLMTAFDSVQNLVVGSGGTINGNGQVW 181  
97 LKNIDLEVLTITLFTNDTDYQANSFKQGFQNAITTFQLOG-EDVNMVGGGTINGNGQVW 155  
182 WPSCKINKSLPRDAPTLTLFWNCKNKKVNNKSKNAQOIHKFESCTNVVNSNLMINA 241  
156 YDIYAEDDLTL---PRLMGITGLNGGTIGPKLRISPDYHFFVANSNVLPFGIDISG 211  
242 SAKS-----PNTDGVHVSNTQYIQISDTIIGTGDICISYSGSQNVQATNTTCGPGHGISI 297  
212 YKSDNDKAKNTDMDYRSNNIVIQNSVINNGDDCVSFKNSNIIIVQNLHCHNGSHGISV 271  
298 GSISGSGNSE-----AVSNVTVNEAKITGAENGV-----RIKTWQ----- 332  
272 GSIGQYKDEVDIYENYVYVNIISMFNASVCINFNHIIIDFLTLWLQDMARIKIVWPGTSPALS 341  
333 -----GGSGQANIKPLINWEMODVKKPILIIDQNYCR--VEPCIQGSAVQVKNVYENI 385  
332 ADIQGGGSGSVKNTIYDVALIDNVDALEITQCYGKNTLLCNEYFSSSLTISDVHITKNF 391  
386 KGTSA-----TKVAIKFCSTNPFCEGIIIMENINLVGSESGKPESEATCKNVHFNNAEH 437  
392 RGTTSSEDEPYVGTIV-----GSSPDTCSDIYTSNINVTSPDG--TNDVFCDDV---DESL 442  
438 VTPHCTIS 444  
:::|::|:

Db 443 LSVNCTA 449

RESULT 9  
US-08-780-869-5  
Sequence 5, Application US/08780869  
Patent No. 5830737  
GENERAL INFORMATION:  
APPLICANT: KUSTERS-VAN SOMEREN, MARGO A.  
APPLICANT: MULLER, YVONNE  
APPLICANT: KESTER, HERMANUS C.M.  
APPLICANT: VISSER, JACOB  
APPLICANT: VAN COYEN, ALBERT J.J.  
APPLICANT: ROLIN, CLAUD  
TITLE OF INVENTION: CLONING AND EXPRESSION OF THE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/780,869  
FILING DATE: 24-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/290,978  
FILING DATE: 17-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 4615-0044.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 452 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-780-869-5

Query Match 12.5%; Score 298.5; DB 2; Length 452;  
Best Local Similarity 25.1%; Pred. No. 1.3e-18;  
Matches 107; Conservative 73; Mismatches 166; Indels 81; Gaps 16;

74 KVINLVSFGAKDGKTYDNIAPFOANNEACSSRTPVQFVVPKKNYLL--KQITFGSPC 130  
48 KTCVASHGEGTDDSYIISALNQ-----CNHGKVF--DEKKEYIIGALNMTF---- 96  
131 RSSISVKIFGSL-----EASSKISDYKDRRLMTAFDSVQNLVVGSGGTINGNGQVW 181  
97 LKNIDLEVLTITLFTNDTDYQANSFKQGFQNAITTFQLOG-EDVNMVGGGTINGNGQVW 155  
182 WPSCKINKSLPRDAPTLTLFWNCKNKKVNNKSKNAQOIHKFESCTNVVNSNLMINA 241  
156 YDIYAEDDLTL---PRLMGITGLNGGTIGPKLRISPDYHFFVANSNVLPFGIDISG 211  
242 SAKS-----PNTDGVHVSNTQYIQISDTIIGTGDICISYSGSQNVQATNTTCGPGHGISI 297  
212 YKSDNDKAKNTDMDYRSNNIVIQNSVINNGDDCVSFKNSNIIIVQNLHCHNGSHGISV 271  
298 GSISGSGNSE-----AVSNVTVNEAKITGAENGV-----RIKTWQ----- 332



```
/ FILING DATE: 14 MAY 1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KOKULIS, PAUL N.
/ REGISTRATION NUMBER: 16773
/ REFERENCE/DOCKET NUMBER: 202390/R 7262 (V)
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 861-3000
/ TELEFAX: (202) 822-0944
/ TELEX: 6714627 CUSH
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 440 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-061-062A-6

Query Match      10.1%; Score 241; DB 1; Length 440;
Best Local Similarity 24.0%; Pred. No. 2e-13;
Matches 88; Conservative 61; Mismatches 186; Indels 32; Gaps 11;

QY 70 KNGIKVINLSPGAKGDKTYNDIAFEQAWNEACSRTPVQFVPPKKNKYLKQITFSGP 129
Db 33 KGATKCNILSYGAVADNSTDVGPALTSAM-AACKSGGLV--YIPSGNYALNTWVTLTG 89
QY 130 CRRSISVK--IFSGLEASSKISDYKDRRLMIAFDSYQNLVGGGGTNGQVWMPSSCK 187
Db 90 SATALDGLDIYRTGASGNMLAVDTDFELFSSISK-----GAVQGGVYVHAEGTY 143
QY 188 INKSLPCRDAPALTLFWNCKNLKVNLSKNAQOIHKEBSCNVAASNLMTNASKSPN 247
Db 144 GARILRLTDV-----THFSVDVILVDAAPAFHFTMDTCSDEGVYMMALRGNEG-G 193
QY 248 TGVVHNSNTOYQISPTIITGDDCISIVSGSONVATNITGPRGHGISIGISGNSSEA 307
Db 194 LDGIDVWSN-IWHDVEVTNKECVTKSPANNILVESIYCNWSSGCGAMSG--ADT 249
QY 308 YVSNVTVNEAKLIGENGVRIRKTWQSGSQASNIKELNVEMQVKKYPIIIDQNYCDRVEP 367
Db 250 DVTDIYRVNVYTWSSNQWMIKS--NGSGTGVSNVLENIIGHNAYSLLIDGYWSMTAV 308
QY 368 CIQGSNAVQKVVYENIKGT--SATKVAIKFDCSTNPPCGEIMENINLVESGKPSK 424
Db 309 A--GDGVQLNINITVKNWKGTEANGATRPPIRVVCSDTAPCTDLTDIAIWTSGSSEL 365
QY 425 ATCKNVH 431
Db 366 YLCRSAY 372

RESULT 13
US-08-061-062A-8
/ Sequence 8, Application US/08061062A
/ Patent No. 5550045
/ GENERAL INFORMATION:
/ APPLICANT: MUSTERS, WOUTER
/ APPLICANT: STAM, HEIN
/ APPLICANT: SUYKERBOEK, MARIA E.
/ APPLICANT: VISSER, JACOB
/ APPLICANT: VERBAKEU, Johannes M.
/ TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
/ TITLE OF INVENTION: ENCODING A RIPENING FORM OF A POLYPEPTIDE HAVING
/ TITLE OF INVENTION: RHANOGALACTURONASE ACTIVITY
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CUSHMAN DABRY & CUSHMAN
/ STREET: 1100 NEW YORK AVENUE, N.W.
/ CITY: WASHINGTON, D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-3918
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Disk
```

```
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/061,062A
/ FILING DATE: 14 MAY 1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KOKULIS, PAUL N.
/ REGISTRATION NUMBER: 16773
/ REFERENCE/DOCKET NUMBER: 202390/R 7262 (V)
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 861-3000
/ TELEFAX: (202) 822-0944
/ TELEX: 6714627 CUSH
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 440 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-061-062A-8

Query Match      10.1%; Score 241; DB 1; Length 440;
Best Local Similarity 24.0%; Pred. No. 2e-13;
Matches 88; Conservative 61; Mismatches 186; Indels 32; Gaps 11;

QY 70 KNGIKVINLSPGAKGDKTYNDIAFEQAWNEACSRTPVQFVPPKKNKYLKQITFSGP 129
Db 33 KGATKCNILSYGAVADNSTDVGPALTSAM-AACKSGGLV--YIPSGNYALNTWVTLTG 89
QY 130 CRRSISVK--IFSGLEASSKISDYKDRRLMIAFDSYQNLVGGGGTNGQVWMPSSCK 187
Db 90 SATALDGLDIYRTGASGNMLAVDTDFELFSSISK-----GAVQGGVYVHAEGTY 143
QY 188 INKSLPCRDAPALTLFWNCKNLKVNLSKNAQOIHKEBSCNVAASNLMTNASKSPN 247
Db 144 GARILRLTDV-----THFSVDVILVDAAPAFHFTMDTCSDEGVYMMALRGNEG-G 193
QY 248 TGVVHNSNTOYQISPTIITGDDCISIVSGSONVATNITGPRGHGISIGISGNSSEA 307
Db 194 LDGIDVWSN-IWHDVEVTNKECVTKSPANNILVESIYCNWSSGCGAMSG--ADT 249
QY 308 YVSNVTVNEAKLIGENGVRIRKTWQSGSQASNIKELNVEMQVKKYPIIIDQNYCDRVEP 367
Db 250 DVTDIYRVNVYTWSSNQWMIKS--NGSGTGVSNVLENIIGHNAYSLLIDGYWSMTAV 308
QY 368 CIQGSNAVQKVVYENIKGT--SATKVAIKFDCSTNPPCGEIMENINLVESGKPSK 424
Db 309 A--GDGVQLNINITVKNWKGTEANGATRPPIRVVCSDTAPCTDLTDIAIWTSGSSEL 365
QY 425 ATCKNVH 431
Db 366 YLCRSAY 372

RESULT 14
US-08-536-150-6
/ Sequence 6, Application US/08536150
/ Patent No. 6013489
/ GENERAL INFORMATION:
/ APPLICANT: MUSTERS, WOUTER
/ APPLICANT: STAM, HEIN
/ APPLICANT: SUYKERBOEK, MARIA E.
/ APPLICANT: VISSER, JACOB
/ APPLICANT: VERBAKEU, Johannes M.
/ TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
/ TITLE OF INVENTION: ENCODING A RIPENING FORM OF A POLYPEPTIDE HAVING
/ TITLE OF INVENTION: RHANOGALACTURONASE ACTIVITY
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CUSHMAN DABRY & CUSHMAN
/ STREET: 1100 NEW YORK AVENUE, N.W.
```

CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/536,150  
FILING DATE: 29-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/061,062  
FILING DATE: 14 MAY 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 202390/R 7262 (V)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 440 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-536-150-6

Query Match 10.1%; Score 241; DB 3; Length 440;  
Best Local Similarity 24.0%; Pred. No. 2e-13;  
Matches 88; Conservative 61; Mismatches 166; Indels 32; Gaps 11;

QY 70 KNGIKVINIVSFGAKGDKTYDNIAPQAMNEACSSRTPVQFVVPKKNYLLKQITFSGP 129  
DB 33 KGATKTCNIIISYGAVANSTDVGPALTSAM-AAKSGGLV--YIPSGNYALNTWVTLTG 89  
QY 130 CRSSISVK--IFGSLKSSKISDYKDRRLWIAFDSVQNLVVGSGGTINGNQVWMPSSCK 187  
DB 90 SATAIQLDGIILYRTGASGNMIAVLTDTTDFELFSTSK-----GAVQGFYVYHAEQTY 143  
QY 188 INKSLPCRDAPALTFNCKLKNLKSNAQOIHIFKFSCTNVVASNLMINASAKSPN 247  
DB 144 GARILRLTDV-----THFSVHDVILVDAFHFMTDTCSDGEVYNNALIRGNGE-G 193  
QY 248 TDGVHVSNTQYIQISDTIIGTDDCISIVSGSQVQATNITCGPHGISIGSLGSGNSEA 307  
DB 194 LDGIDVWGSN-IWHDVEVTNKDECVTVKSPANNILVESIYCNWGGCAMSGLG---ADT 249  
QY 308 YVSNVTNNEAKITGAENGVRKIKTWGSGSQASNIKEFLNEMQDVKYPIIIDQNYCDRVEP 367  
DB 250 DVTDIVRNVYTWSSNQMYIKS-NGSGTVSNVLLNFIGHGNAYSLDIDGYWSSMTAV 308  
QY 368 CIOQFSAVQKNVYVENIKGT---SATKVAIKPDCSTNFPCEGIIMENINLVGSGKPS 424  
DB 309 A---GDGVQNLNITVKKWKGTENGATRPPIRVVCSPTAPCTDLTLDIAIWTSSGSEL 365  
QY 425 ATCKNVH 431  
DB 366 YLCRSAY 372

RESULT 15  
US-08-536-150-8  
Sequence 8, Application US/08536150  
Patent No. 6013489  
GENERAL INFORMATION:  
APPLICANT: MUSTERS, WOUTER  
APPLICANT: STAM, HEIN  
APPLICANT: SUYKERBOYK, MARIA E.  
APPLICANT: VISSER, JACOB

APPLICANT: VERBAKEL, Johannes M.  
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA  
TITLE OF INVENTION: ENCODING A RIPENING FORM OF A POLYPEPTIDE HAVING  
TITLE OF INVENTION: RHANNOGALACTURONASE ACTIVITY  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABRY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/536,150  
FILING DATE: 29-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/061,062  
FILING DATE: 14 MAY 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 202390/R 7262 (V)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 440 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-536-150-8

Query Match 10.1%; Score 241; DB 3; Length 440;  
Best Local Similarity 24.0%; Pred. No. 2e-13;  
Matches 88; Conservative 61; Mismatches 166; Indels 32; Gaps 11;

QY 70 KNGIKVINIVSFGAKGDKTYDNIAPQAMNEACSSRTPVQFVVPKKNYLLKQITFSGP 129  
DB 33 KGATKTCNIIISYGAVANSTDVGPALTSAM-AAKSGGLV--YIPSGNYALNTWVTLTG 89  
QY 130 CRSSISVK--IFGSLKSSKISDYKDRRLWIAFDSVQNLVVGSGGTINGNQVWMPSSCK 187  
DB 90 SATAIQLDGIILYRTGASGNMIAVLTDTTDFELFSTSK-----GAVQGFYVYHAEQTY 143  
QY 188 INKSLPCRDAPALTFNCKLKNLKSNAQOIHIFKFSCTNVVASNLMINASAKSPN 247  
DB 144 GARILRLTDV-----THFSVHDVILVDAFHFMTDTCSDGEVYNNALIRGNGE-G 193  
QY 248 TDGVHVSNTQYIQISDTIIGTDDCISIVSGSQVQATNITCGPHGISIGSLGSGNSEA 307  
DB 194 LDGIDVWGSN-IWHDVEVTNKDECVTVKSPANNILVESIYCNWGGCAMSGLG---ADT 249  
QY 308 YVSNVTNNEAKITGAENGVRKIKTWGSGSQASNIKEFLNEMQDVKYPIIIDQNYCDRVEP 367  
DB 250 DVTDIVRNVYTWSSNQMYIKS-NGSGTVSNVLLNFIGHGNAYSLDIDGYWSSMTAV 308  
QY 368 CIOQFSAVQKNVYVENIKGT---SATKVAIKPDCSTNFPCEGIIMENINLVGSGKPS 424  
DB 309 A---GDGVQNLNITVKKWKGTENGATRPPIRVVCSPTAPCTDLTLDIAIWTSSGSEL 365  
QY 425 ATCKNVH 431  
DB 366 YLCRSAY 372

Search completed: July 21, 2004, 17:09:27

✓ Tue Jul 27 08:55:41 2004

us-10-018-604-2.ra1

Page 9

Job time : 29.379 secs

---

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

## OM protein - protein search, using sw model

Run on: July 21, 2004, 17:08:27 ; Search time 79.2655 Seconds

(without alignments)  
1802.048 Million cell updates/sec

Title: US-10-018-604-2

Perfect score: 2390  
Sequence: 1 MVIQNSILLILIIIFASSIS.....VTPHCTSEISEDEALVNY 457Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285356 seqs, 312560742 residues

Total number of hits satisfying chosen parameters: 1285356

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA.\*

```
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1113	46.6	433	13	US-10-151-668-2
2	1011	42.3	393	16	US-10-437-963-127090
3	1010	42.3	452	16	US-10-437-963-127091
4	988.5	41.4	444	16	US-10-437-963-111044
5	870	36.4	319	12	US-10-425-114-69361
6	808	33.8	508	16	US-10-437-963-154485
7	787.5	32.9	514	10	US-09-847-208-69
8	785	32.8	534	16	US-10-437-963-134257
9	739	30.9	485	16	US-10-437-963-175508
10	734	30.7	503	16	US-10-437-963-172689
11	728	30.5	443	12	US-10-424-599-235443
12	728	30.5	456	12	US-10-425-114-44707
13	724.5	30.3	398	12	US-10-362-091-4
14	723.5	30.3	492	12	US-10-424-599-284649
15	701.5	29.4	400	16	US-10-437-963-155040

16	700	29.3	573	12	US-10-425-114-43413	Sequence 43413, A
17	699	29.2	496	12	US-10-425-114-66151	Sequence 66151, A
18	682	28.5	571	16	US-10-437-963-159100	Sequence 159100, A
19	681	28.5	737	16	US-10-437-963-162350	Sequence 162350, A
20	678	28.4	457	16	US-10-437-963-146166	Sequence 146166, A
21	666	27.9	393	16	US-10-437-963-116114	Sequence 116114, A
22	664.5	27.8	415	12	US-10-424-599-238512	Sequence 238512, A
23	660	27.6	422	12	US-10-424-599-176330	Sequence 176330, A
24	653.5	27.3	385	12	US-10-425-114-41984	Sequence 41984, A
25	653.5	27.3	411	12	US-10-425-114-38472	Sequence 38472, A
26	653.5	27.3	445	12	US-10-425-114-51531	Sequence 51531, A
27	649.5	27.2	449	16	US-10-437-963-110174	Sequence 110174, A
28	643.5	26.9	602	16	US-10-437-963-127089	Sequence 127089, A
29	642	26.9	320	12	US-10-437-963-204806	Sequence 204806, A
30	641.5	26.8	544	16	US-10-427-963-110126	Sequence 110126, A
31	635.5	26.6	539	12	US-10-424-599-176466	Sequence 176466, A
32	632	26.4	386	12	US-10-425-114-50991	Sequence 50991, A
33	630	26.4	410	12	US-10-425-114-38475	Sequence 38475, A
34	630	26.4	414	12	US-10-425-114-62958	Sequence 62958, A
35	630	26.4	414	12	US-10-425-114-64531	Sequence 64531, A
36	627	26.2	407	12	US-10-425-114-65565	Sequence 65565, A
37	627	26.2	410	12	US-10-425-114-46486	Sequence 46486, A
38	627	26.2	412	12	US-10-425-114-46801	Sequence 46801, A
39	627	26.2	413	12	US-10-425-114-52982	Sequence 52982, A
40	627	26.2	413	12	US-10-425-114-56708	Sequence 56708, A
41	627	26.2	415	12	US-10-425-114-46800	Sequence 46800, A
42	627	26.2	415	12	US-10-425-114-54673	Sequence 54673, A
43	627	26.2	415	12	US-10-425-114-64510	Sequence 64510, A
44	627	26.2	416	12	US-10-425-114-65564	Sequence 65564, A
45	627	26.2	416	12	US-10-425-114-72785	Sequence 72785, A

## ALIGNMENTS

```
RESULT 1
US-10-151-668-2
; Sequence 2, Application US/10151668
; Publication No. US2002018460A1
; GENERAL INFORMATION:
; APPLICANT: UVSROV, Peter
; APPLICANT: CHILD, Robin
; APPLICANT: VAN ONCKELIN, Henri
; APPLICANT: PRINSEN, Els
; APPLICANT: BORKHARDT, Bernard
; APPLICANT: SANDER, Lilli
; APPLICANT: PETERSEN, Morten
; APPLICANT: BUNDGARD, Poulsen, Gert
; APPLICANT: BOTTERMAN, Johan
; TITLE OF INVENTION: Seed Shattering
; FILE REFERENCE: 2121-0138P
; CURRENT APPLICATION NUMBER: US/10/151,668
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US/09/051,239
; PRIOR FILING DATE: 1998-09-28
; PRIOR APPLICATION NUMBER: PCT/EP96/04313
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: EP 95 402241.4
; PRIOR FILING DATE: 1995-10-06
; PRIOR APPLICATION NUMBER: EP 95 203328.0
; PRIOR FILING DATE: 1995-12-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Brassica napus
; FEATURE:
; OTHER INFORMATION: Strain cv. Topaz.
US-10-151-668-2
Query Match 46.6%; Score 1113; DB 13; Length 433;
Best Local Similarity 49.9%; Pred. No. 1.8e-97;
```



```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69361
LENGTH: 319
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73002C03_FLI.pep
US-10-425-114-69361

Query Match          36.4%; Score 870; DB 12; Length 319;
Best Local Similarity 50.8%; Pred. No. 2.le-74;
Matches 162; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY      132 SSIISVKEFGSLSEASKSISDY--KORRLIMAFDSYQNIVVGGGGTINGNGVWVPSSCKIN 189
        |:::|::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1 SSVTLLTKGTLIVASPFRADMSDNDRRRMYYFRSIDKLITVNGGAIDNGEHWHSCKIN 60
        |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      190 KSLPCRDAPALATFWNCKNLKVNNLKSKNAQQIHKEESCCTNVVASNMIMINASAKSPNTD 249
        |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       61 KALPCKREAPTALSFHYCVLDKVEDKLIVNSQGIHMSVEDESNALLARLSTIAPGRSPTTD 120
        |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      250 GVHSVNTQYIQISTITIGTDDCLSIYSGSQNOVAINTTCGPBGHGISTGISGSNSERYV 309
        |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

QY      310  SNVTVNFAKIIIGANGVRITKWTQSGQASNIKFLYNEMOPVKYPIIIDQYCDRVBECI 369
      181  SGITISVQILHGTINGRIRIKTYQSSGYARDITQNMWMDVKNPIIIDQYCDKAKCG 240
Db
QY      370  QQPSAVQKVVNENIKGTSATKYAIKFDGCTNFCGIIIMENINLVGSESK-DSEATCK 428
      241  EQESAQVQGVFKNRINIGTSTKDAIKMNCSENVPCGILTQNIIDLKMQDGKNTSPCTCG 300
Db
QY      429  NVHFNMAEHVPH-CTSL E 446
      301  NAKWTEFGTVRPPDCTAIK 319
Db

RESULT 6
US-10-437-963-154485
; Sequence 154485, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21153221/B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 154485
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54340C.1.pep
US-10-437-963-154485

Query Match      33.8%; Score 808; DB 16; Length 508;
Best local similarity 36.4%; Pred. No. 3,66-68;
Matches 173; Conservative 91; Mismatches 167; Indels 44; Gaps 7

```

```

QY      8 ILLIIIFASSISTCRSNVIDNLFKQVYVNIILBOFAHDFOAYL-----SY 54
Db      18 VLVALMTYIMAAKTASGGGGDGGAGGHD-----HDHEGFKLKMGNGADAKDY 69
QY      55 LSKNIESNNIDKVDK-----GIKIVNLSFGAKDGTNYNIAEQ 97
Db      70 LNMDDDDDDDEDEEEBAEQVMAAKCRPPAGRVNVNDSFGAAGDGSDDTEALIN 129
QY      98 AMNACSSRTVQVVPKKNKYLLKQITFSGPCRSSISVKTFGSLFASKISDY--KRR 155
Db      130 AMKPACSIINNAV-FLVPGGRRYKVGAAFIQPCKNRMITIQGTTVADESEMDPASPR 188
QY      156 LMIAPDSYQNLVVGGGGTINGNGQVWMPSCSKINKSLPCRDAPALTTPWCKNLKVNLIK 215
Db      189 LMLFSLGAGARIQGGGLIDGSGSKWMAKNSCKIDRSKCKAPALTITDSGRGVSRLR 248
QY      216 SKNAQOIHKEESCCTNVVNASLMLTINASAKSPNTDGVHVSNTQYIQISDTITGDDCISI 275
Db      249 LQNAQOMELTVSRSDVRLASVRVDSPEDSPTDGIHVADSTAVTIQSCRITATGDDCISI 308
QY      276 VSGSQVQATNITGPGHGISIGSGNSSEAYSVNTVNAKTIIGENGRIKTMOGGS 335
Db      309 SNGSFAVRMRDIDGPGHGISIGSGGGAFAVDGSLDGAARVAQNGVRITWOGGA 368
QY      336 GQASNIKFLNVEMQDVKYPITIDONYCDRVEPCIQOFSAVQKVVYENIKGTSATKVAI 395
Db      369 GYVRNVAFAGRVGVDPHPIVIDQFCATPCRNRTSNVAVSGVFENIIGTARRAEAI 428
QY      396 KFDCTNPPCEGIIEMENINLVGSEKSEATCKN--VHFNNAEHVTHCHTSLEIS 448
Db      429 RLACSDAVPCVGIIVLSDIDLEREDGGGVQTVNCAMGFDDG-RVSPADCLRTS 482

```

## RESULT 7

```

US-09-847-208-69
; Sequence 69, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847.208
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Cryptomeria japonica (Japanese cedar)
US-09-847-208-69

```

```

Query Match      32.9%; Score 787.5; DB 10; Length 514;
Best Local Similarity 38.5%; Pred. No. 3,3e-66;
Matches 173; Conservative 78; Mismatches 163; Indels 35; Gaps 9;

QY      8 ILLIIIFASSISTCRSNVIDNLFKQVYVNIILBOFAHDFOAYLSTYSKNTESNNIDK 67
Db      12 VAMOLIIIMAA-----EDQSAQIMLDSIDQ-----YL-----RSNRSLRK 47
QY      68 VD---KNGIKIVNLSFGAKDGTNYNIAEQAMNEACSSRTVQVFPVKKNKYLLKQI 124
Db      48 VESRRHAINIFNEVEKGAVDGKHDCTEASTAMQAC-KKPSAMLLVPSNKKFVVNLT 106
QY      125 TFGSPCRSSISVKTFGSLHSAASKISDYKDRRLMTAFDSVQNLVVGGGGTINGNGQVWMP 184
Db      107 FPNGPQPHFTFKYDGIILAAVQNPASMKNNRIMLQFAKLTFGLMGKVIDGQGRKWWAG 166
QY      185 SCR-INKSLPC--RDAPALTTPWCKNLKVNLIKSKNAQOIHKEESCCTNVVNASLMLTINA 241
Db      167 QCKWVNGREICNDRPETAIKFDSFTGLIIGLKLXNNSPEPHLVFGNCEGVKIIIGISITA 226

```

```

QY      242 SAKSPNTDGVHVSNTQYIQISDTITGTDPCISIVSSQNVQVATNITGPGHGISISLG 301
Db      227 PRDSPNTDIDIPASKKNHLPKNTGTTGDDCVALTGSSNITVIEDLIGPGHGISISLG 286
QY      302 SGNSEAYSVNTVNAEKIIGAENGRIKTMOGSGQASNIKFLNVEMQDVKYPITIDONY 361
Db      287 RENBRAEVSYVHVAGAKRIDQNGRLIKTMOGSGMAHIIYENEMINSENPILINQFY 346
QY      362 CDRVEPCIQOFSANQVKNVYENIKGTSATKVAIKFDCTNPPCEGIIEMENINLVGSEK 421
Db      347 CTSASACQMSASAVQIDQVYKNIRGTSATAAIQLKCSDEMPCKDILSLKLTSGK 406
QY      422 PSBATCKNVHFN--AEHVTHCHTSLEIS 448
Db      407 --IASCLNDNANGYFSGHVTPACKNLSFS 433

```

## RESULT 8

```

US-10-437-963-134257
; Sequence 134257, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 134257
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36049C.1.pep
US-10-437-963-134257

```

```

Query Match      32.8%; Score 785; DB 16; Length 534;
Best Local Similarity 41.1%; Pred. No. 6.1e-66;
Matches 157; Conservative 57; Mismatches 146; Indels 22; Gaps 4;

QY      81 FGAAGDGTNYNIAEQAMNEACSSRTVQVFPVKKNKYLLKQITFSGPCRSSISVKTFG 140
Db      130 FGAVDGTINDTQAFPLDMMAAKASTEPAYLAIVAGKTYQVWPRLAOPCKKKLKLMSLG 189
QY      141 SLEASKISDY--KDRRLMTAFDSVQNLVVGGGGTINGNGQVWMPSCSKINKSLPCR--D 196
Db      190 TIAPASPDWAGBDPKMLIVFRVDDLVSGGGTIDMGAEWMAWARSCKRKXPCSTVS 249
QY      197 APTALTTPWCKNLKVNLIKSKNAQOIHKEESCCTNVVNASLMLTINASAKSPNTDGVHVSNT 256
Db      250 APKALQEEBCRRVAVQITWONGPQFHLMTFRCDVNASFLRVVAPSSPTDGIHLMDT 309
QY      257 QYIQISDTITGTDPCISIVSSQNVQVATNITGPGHGISIGSGNSSEAYSVNTVNAE 316
Db      310 THAQIMNLTSTGDCYSWANCSDVAKYDSCGPGHGISIGSGKRTTRINENRVDT 369
QY      317 AKIIGAENGRIKTMOGSGQASNIKFLNVEMQDVKYPITIDONYCDRVEPCIQO----- 371
Db      370 CLLTNTNGVRIKMGQMGYAHNLREFGIYMKVNSPILIIDYQCDQPTPCANQSTSS 429
QY      372 -----FS-----AVQVKNVYENIKGTSATKVAIKFDCTNPPCEGIIEMENINLVGE 418
Db      430 SLKLGNSSIMQGTQAVAEVRRIEFGAGIRGTSATQAIKLACSDAVPCPDELRLRVNLTMV 489

```

```
QY 419 SGPSEATCKNNHNNAEHVT 440
DB 490 GGGAGAGFCHRASGKAAGAVP 511

RESULT 9
US-10-437-963-175508
; Sequence 175508, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 175508
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73346C.1.pep
US-10-437-963-175508

Query Match 30.9%; Score 739; DB 16; Length 485;
Best Local Similarity 42.6%; Pred. No. 1.3e-61;
Matches 153; Conservative 62; Mismatches 128; Indels 16; Gaps 7;

QY 75 VINVSFGAKGDKTYDNIAFEQANNACSSRTPVQFVVPKKNYLLKQITSGPCRSSI 134
DB 89 VEFVRAAGVGDGTDTDTTEFAAMPAAACAVESAV-LSVPSDGTFTTTTTPGCKPGI 147
QY 135 SVKIFPSL--EASSKISDYKPRRLMTAFDSVQNLVVGSGGTINGQVWPPSSCKINKS 191
DB 148 VFEQVDVLMPPDPDCMPBSNRRQMLVFESNLDGLTRGAGTLEGGEQWMLPCKPHRG 207
QY 192 L-----PCRDAPTLTFMNCNKLKNNLKSNAQOIHIFESCTNVASNMIMINASAK 244
DB 208 PNCSTLRGPC-DSPTLVRPFMSRNLVVEGLRVENSPEFHRFDGCSVRVDGLSISSPAN 266
QY 245 SPNTDGVHVSNTQYIISDTIIGTDDCISIVSGSQNVQATNITCGPHGISIGLSGSGN 304
DB 267 SPNTDGHVHNTQRAVIAVNSMISNGDDCISIGTSYVDVIDQNVSCGPHGISIGLSGVHN 326
QY 305 SEAYVSNVYVNEAKITIGANGVAKITWQSSGQASINIKFLNVEMQDYKYIITIDQNYC-D 363
DB 327 SQACVAVNIVRNVRNIRNSDGLRIKTWQGGSGSVSGINFTVSMENVRNCIITIDQYYCUD 386
QY 364 RVSPCICQGFSAVQVNVYENIKGTSATKVA-IFKDCSTNFPCEGIMENINLVGSGK 421
DB 387 K-RCMQNSTAVHTVTSYANVAGSYVRAAPHFACSDVPCNTITMSVELLPSSGE 443

RESULT 10
US-10-437-963-172689
; Sequence 172689, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172689
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70800C.1.pep
US-10-437-963-172689

Query Match 30.7%; Score 734; DB 16; Length 503;
Best Local Similarity 40.7%; Pred. No. 4.2e-61;
Matches 153; Conservative 66; Mismatches 147; Indels 10; Gaps 6;

QY 78 VLSFGAKGDKTYDNIAFEQANNACSSRTPVQFVVPKKNYLLKQITSGPCRSSISVK 137
DB 114 VDFGAGAGGVTDDBDALTAMDTACADGAGVLLAAGRSSFLHTTVFTGPGCGSVTLQ 173
QY 138 IFGSLFASSKISDY-KDRRLMTAFDSVQNLVVGSGGTINGQVWPPSSCKINKS---- 191
DB 174 VDGTVAPSEPAWPNANRRNMLVFYRADGVSIVGAGLIDGKQKXWDLPCPKHKGANTH 233
QY 192 LPCRDAPTLTFMNCNKLKNNLKSNAQOIHIFESCTNVASNMIMINASAKSPNTDGV 251
DB 234 GPC-DSPVAMRAFRASNNVVRGLKVONSPEFHRFPCNCGVRVDGLSISPDLSPTDGI 292
QY 252 HVSNTQYIISDTIIGTDDCISIVSGSQNVQATNITCGPHGISIGLSGNSSEAYVSN 311
DB 293 HVNTSDVLTITNVVSGNDQCVSIGAGTLNVHIENTVCGPHGISIGLSGKAGTKACVAN 352
QY 312 VTYNEAKITIGANGVAKITWQSSGQASINIKFLNVEMQDYKYIITIDQNCDEVEPTIQ 371
DB 353 VTYRNVIHNSDNGVRIKTWQSSGSVAFAFENVMDAARNPITIDQYVC-LSKSGEN 411
QY 372 FSAVQKVNYYENIKGTSATK-VAIFKDCSTNFPCEGIMENINLVGSGKP-SEATCKN 429
DB 412 TVAVFVNGSVYSIGRTGYVRGPHIFGCSDAVPCNITLSDVELLPASGDYDDEPCWN 471
QY 430 VEPNNAEHVTPHCTSL 445
DB 472 VVGNAATPTVPVPSCL 487

RESULT 11
US-10-424-599-235443
; Sequence 235443, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 235443
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_54632C.1.pep
US-10-424-599-235443

Query Match 30.5%; Score 728; DB 12; Length 443;
Best Local Similarity 42.9%; Pred. No. 1.3e-60;
```

Matches 158; Conservative 51; Mismatches 139; Indels 20; Gaps 10;

QY 77 NVLSFGAKGDKTYDNIAFEQAMNEACSSRTVPQVFPKNNKYLKQITFSGP-CRSSIS 135  
Db 47 NVLDYGAAGDGHADDTKAFEDMAAACVKEGST-MVPSGSVFLKPISTFSGENCEPIIV 105  
QY 136 VKIFGSLFASAKISIDYKDRRL-WIAFDSVQNLVVGSGGTINGNGVWV-----PSSC 186  
Db 106 FOLDGKIAPTSSEAWSGTLQWLEFSKLNITITIRGKVIDQGSVWNNNSPTYNPLEV 165  
QY 187 KI--NKSIPCRDAPALTFMNCNKLKVNLSKNAQOIHIFESCTNVVANSNLINSAK 244  
Db 166 MLESNGRPL-STKPTALRFYSDDGVTITGTONSQQTHLKFDSCNTNVQVSGISVSSGD 224  
QY 245 SPNTDGVHVSNTQYIQISDTITIGTDCISIVSGSQNVQATNITGPGHGISIGSLGSGN 304  
Db 225 SPNTDGHLOQNSQNVVYSSITLACDDCVSIQTGSDIYHNVNCPGPHGISIGSLGREN 284  
QY 305 SEAYVSVNTVNEAKITGAENGVRITKMGSGSQASNIKFLVEMQDVYKPIIIDQNYCDR 364  
Db 285 TKACVRNVTVDVITQNTLTGVRITKMGSGSQVQINIMFSNVQVSGVQPTSIDQYCDG 344  
QY 365 VEPICQFSAVQVKNVVENIKGTSATKVAIKFDSCSTNFPCEGIMENINLVSGSGKPE 424  
Db 345 GR-CRNESAAVAVSGIHVNVKGT-YTKPTIFYACSDMLPCSGITLDTIQL--BSAQ--E 398  
QY 425 ATCKNVHP 432  
Db 399 TKNSNVPR 406

RESULT 12  
US-10-425-114-44707  
; Sequence 44707, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 44707  
; LENGTH: 456  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700847226\_FLI.rep  
US-10-425-114-44707

Query Match 30.5%; Score 728; DB 12; Length 456;  
Best Local Similarity 42.9%; Pred. No. 1.4e-60;  
Matches 158; Conservative 51; Mismatches 139; Indels 20; Gaps 10;

Db 228 SPNTDGHLOQNSQNVVYSSITLACDDCVSIQTGSDIYHNVNCPGPHGISIGSLGREN 297  
QY 305 SEAYVSVNTVNEAKITGAENGVRITKMGSGSQASNIKFLVEMQDVYKPIIIDQNYCDR 364  
Db 298 TKACVRNVTVDVITQNTLTGVRITKMGSGSQVQINIMFSNVQVSGVQPTSIDQYCDG 357  
QY 365 VEPICQFSAVQVKNVVENIKGTSATKVAIKFDSCSTNFPCEGIMENINLVSGSGKPE 424  
Db 358 GR-CRNESAAVAVSGIHVNVKGT-YTKPTIFYACSDMLPCSGITLDTIQL--BSAQ--E 411  
QY 425 ATCKNVHP 432  
Db 412 TKNSNVPR 419

RESULT 13  
US-10-362-091-4  
; Sequence 4, Application US/10362091  
; Publication No. US20040049809A1  
; GENERAL INFORMATION:  
; APPLICANT: Instituto de Ciencia Aplicada e Tecnologia (IGAT)  
; TITLE OF INVENTION: Pear genes codifying for b-galactosidase, Pectin Methyltransferase,  
; FILE REFERENCE: none  
; CURRENT APPLICATION NUMBER: US/10/362,091  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: PT 102511 C  
; PRIOR FILING DATE: 2000-08-22  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Pyrus communis  
US-10-362-091-4

Query Match 30.3%; Score 724.5; DB 12; Length 398;  
Best Local Similarity 41.4%; Pred. No. 2.4e-60;  
Matches 158; Conservative 64; Mismatches 145; Indels 15; Gaps 10;

QY 64 NIDKVDKNGIKVINVLSPGAKGDKTYDNIAFEQAMNEACSSRTVPQVFPKNNKYLKQ 123  
Db 23 SITVWDAAV-TFSVSSIGAKADGSTDTKAFLSAMSNACASVNPVAYVAGR-FLIGN 80  
QY 124 ITFSGPCR-SSISVKITIGSLFASAKISDYK--DRRLIAPDSVQNLVVGSGGTINGNQ 179  
Db 81 AVFSGPCKNNAITPRLIGTLVAP--SDRYIGNAGNMLFQHVNGVTI-SGGVLDGGGT 136  
QY 180 VWWPSSCKINKSLPCRAPALTATFMNCNKLKVNLSKNAQOIHIFESCTNVVANSNMI 239  
Db 137 GLW--DCK-SGSKCPCPGATLTSNSNNVYVSGLISNGMFIH VVNGCQNVKMGYKV 193  
QY 240 NASAKSPNTDGVHVSNTQYIQISDTITIGTDCISIVSGSQNVQATNITGPGHGISIGS 299  
Db 194 NAAQNSPNTDGIHVQMSGVTIILDSKISTGDDCVSVGGTNTLMIENVACGPHGISIGS 253  
QY 300 LGSQNSZAYVSVNVNEAKITGAENGVRITKMGSGSQASNIKFLVEMQDVYKPIIID 358  
Db 254 LKQDQZBQVQNTVKTVTFTGTENGVRIKMGSPSTGFASILFQHTVMNVQNPVID 313  
QY 359 QNYCDRVEPCIQFSAVQVKNVVENIKGTSATKVAIKFDSCSTNFPCEGIMENINLVGE 418  
Db 314 QNYCPNDKGCQGSQVGVSVTVYQDHLGHSATETVAVKFDCSSNYPNCNGIRLDQVXLT-Y 372  
QY 419 SGKPSBATCKNVHPNNAEHVTP 440  
Db 373 NNQAAEASCIHAGGTAGTAVP 394

RESULT 14  
US-10-424-599-284649  
; Sequence 284649, Application US/10424599

```
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284649
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_39063C.1.pep
US-10-424-599-284649
```

Query Match 30.3%; Score 723.5; DB 12; Length 492;

Best Local Similarity 39.6%; Pred. No. 4.1e-60; Mismatches 157; Indels 15; Gaps 7;

```
Matches 155; Conservative 64; Mismatches 157; Indels 15; Gaps 7;
QY 75 VINVLSEFGAKGDKTYDNTAFEOAMNEACSSRTPVQFVVPKKNYLLKQITPESGCRSS1 134
DB 96 VEDVSEFGAVGDCADDTAFRAAKACAVDSGI-VLAPENVSFKITSTIEGCPCKPG1 154
QY 135 SVKIEGSEASSKISDYKD--RLMLIADSVQNLVVGGGGTNGCQVWMPSCCKIN- 190
DB 155 VQVQDITLVAAPGPNMSPADSENQWLVEYRLDQMLTNGTIEGSGDMWDLPCKPHAG 214
QY 191 -----SLPCRDAPLTALTFWNCKNLKYNLKSNAQOIHIFESCTNVVASNLMINASAK 244
DB 215 PNGKITLSGFC-GSPAMIRRFMSNLVKGLKTONSGQFMHFGCCGVLLDKLISSPKL 273
QY 245 SPNTDGVHVSNTQYIQISDTITGTGDCISIVSGSONVQATNITCGPHGISIGLSGSGN 304
DB 274 SPNTDGIHENSRYKVGCIYNSMISNGDDCISIGPSSNVDIAGLTGCPHSGISIGLSGVHN 333
QY 305 SEAYVSNVTVNNAKTIIGAENGVRITKTWOGSGGASIKLNVEMQVYKPIIIDQNYCR 364
DB 334 SQACVSNLTVRDSIIRRESNGRLIKTWOGMGSSVSLIRENIOEMENVGNCIIIDQYCCLS 393
QY 365 VEPICIQFSAVQKXNVYENIKGTSATKYA-IFDCSTNPPECGLIMENINLVG-ESGKP 422
DB 394 KE-CLNQTSAVHNVDSYNIKGTIVRTAPIHFACSDTVACTNITLSEVELPFEGALL 452
QY 423 SEATCKNVHFNNAEHVTPHCTSLSESEAL 453
DB 453 DDPFCWNAVGTQETLTIPINCLREGDPETV 483
```

## RESULT 15

US-10-437-963-155040

; Sequence 155040, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Bouharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 155040

```
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54842C.1.pep
US-10-437-963-155040
```

Query Match 29.4%; Score 701.5; DB 16; Length 400;

Best Local Similarity 40.1%; Pred. No. 3.9e-58; Mismatches 148; Conservative 69; Mismatches 133; Indels 19; Gaps 8;

```
QY 72 GIKVINVLSEFGAKGDKTYDNTAFEOAMNEACSSRTPVQFVVPKKNYLLKQITPESGCR 131
DB 25 GAAYVNESYDAPDGRDASRALASAWGAACRSEPPATVYVDGE-FVSHSFAFAGPCS 83
QY 132 SS-ISVKIEGSLDEA--SSKISDYKDRRLMLIADSVQNLVVGGGGTNGCQVWMPSCCKIN 169
DB 84 GGRMTVQIDGTLVAPSGYTGSSASSGGEWTVFPHVDGLTV-SGGTLDGRGESLW--ACKAA 140
QY 190 KSLPCRDAPLTALTFWNCKNLKYNLKSNAQOIHIFESCTNVVASNLMINASAKSPNTD 249
DB 141 GHGCGCPDGAITSKVNRSRVVISGVSVASELHVVDGCEGVAQDARLVAPGSSPNTD 200
QY 250 GVHVSNTQYIQISDTITGTGDCISIVSGSONVQATNITCGPHGISIGLSGSGSEAYV 309
DB 201 GIHVOSSSAVTITTGASIGTGDCCISIVGPSTNMLRVHVS CGPHGISIGLSGSEEGGV 260
QY 310 SNVTVNEAKTIIGAENGVRITKTWOGSGGAS-----NIKPLNVEMQDYKPIIIDQNYC 362
DB 261 ENVTVSGAFAVGTENGLRIKTM---GRARSGAYRVGVFEBALMRDVSNNPILIDQSYC 316
QY 363 --DVEPCIQFSAVQKXNVYENIKGTSATKYAIFDCSTNPPECGLIMENINLVGSESG 420
DB 317 PNDGGQGGCHQSSDVQISGVTTTIDQSSASQVAVFDCSASRPGCGLGIDIKLTFDDG 376
QY 421 KPSEATCKN 429
DB 377 KPSEATCKN 385
```

Search completed: July 21, 2004, 17:17:48

Job time : 82.2655 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2004, 17:04:03 ; Search time 29.3576 Seconds  
(without alignments)  
1497.381 Million cell updates/sec

Title: US-10-018-604-2

Perfect score: 2390  
Sequence: 1 MWIQNSILLIILIFASSIS.....VTPHCTSLSEDEALVNY 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2390	100.0	457	2 A25534	polygalacturonase
2	1142	47.8	462	1 S31195	polygalacturonase
3	1128	47.2	461	2 T08215	polygalacturonase
4	1127	47.2	431	2 T46187	polygalacturonase
5	1115.5	46.7	426	2 H84846	probable polygalac
6	1108	46.4	458	2 S71523	polygalacturonase
7	1091	45.6	460	2 T17011	polygalacturonase
8	898.5	37.6	459	2 D96833	hypothetical prote
9	840.5	35.2	468	2 H96728	probable polygalac
10	796.5	33.3	514	2 UC7100	polygalacturonase
11	787.5	32.9	434	2 A96609	probable polygalac
12	787.5	32.9	514	2 S48730	Cry j II protein -
13	787	32.9	514	2 JC2498	second major aller
14	764	32.0	507	2 UC7366	Jun a 2 protein -
15	749.5	31.4	1161	2 B86368	protein F28C11.9 l
16	741.5	31.0	491	2 B86155	probable polygalac
17	736	30.8	423	2 T08213	polygalacturonase
18	728	30.5	405	2 D84871	probable polygalac
19	727	30.4	392	2 S57806	polygalacturonase
20	718.5	30.1	395	2 T05906	probable polygalac
21	718	30.0	435	2 T48638	polygalacturonase
22	715	29.9	392	2 T00669	probable polygalac
23	711.5	29.8	387	2 T04320	polygalacturonase
24	710.5	29.7	365	2 T04319	polygalacturonase
25	708.5	29.6	422	2 S34199	polygalacturonase
26	704.5	29.5	387	2 T04322	polygalacturonase
27	704	29.5	374	2 B85421	probable polygalac
28	704	29.5	374	2 T04672	polygalacturonase
29	702.5	29.4	394	2 G86190	hypothetical prote

30	701.5	29.4	542	2 H86239	protein F20B24.8 l
31	698.5	29.2	452	2 C85024	probable polygalac
32	696	29.1	384	2 E84871	probable polygalac
33	690	28.9	540	2 B96631	probable polygalac
34	688.5	28.8	492	2 C96521	protein F21D18.18
35	687.5	28.8	393	2 S40123	polygalacturonase
36	687	28.7	383	2 T47809	polygalacturonase
37	684	28.6	397	2 F96680	F514.10 [Imported
38	683	28.6	394	2 F86190	hypothetical prote
39	680.5	28.5	394	2 T00668	probable polygalac
40	678.5	28.4	445	2 S34266	polygalacturonase
41	673.5	28.2	387	2 T07591	polygalacturonase
42	673	28.2	444	2 S34200	polygalacturonase
43	667	27.9	407	2 S52006	polygalacturonase
44	662.5	27.7	664	2 A84742	probable polygalac
45	659	27.6	414	2 E85204	polygalacturonase-

## ALIGNMENTS

## RESULT 1

A25534 polygalacturonase (EC 3.2.1.15) precursor - tomato

N:Alternate names: pectinase; poly[1,4-alpha-D-galacturonide]glucanohydrolase  
C:Species: Lycopersicon esculentum (tomato)

C>Date: 30-Jun-1988 #sequence revision 30-Jun-1988 #ext\_change 22-Jun-1999  
C/Accession: A25534; J0156; S06340; S02101; S08504

R:Grierson, D.; Tucker, G.A.; Keen, J.; Ray, J.; Bird, C.R.; Schuch, W.  
Nucleic Acids Res. 14, 8595-8603, 1986

A>Title: Sequencing and identification of a cDNA clone for tomato polygalacturonase.  
A:Reference number: A25534; MUID:87066731; PMID:3786135

A:Accession: A25534

A:Molecule type: mRNA

A:Residues: 1-457 <GRT>

A/Cross-references: GB:X04583; NID:G19291; PIDN:CAA28254.1; PID:G19292

R:DeLapenna, D.; Bennett, A.B.  
Plant Physiol. 86, 1057-1063, 1988

A>Title: In vitro synthesis and processing of tomato fruit polygalacturonase.  
A:Reference number: J0156

A:Accession: J0156

A:Molecule type: mRNA

A:Residues: 1-115 <DEL>

A/Cross-references: GB:M20269; NID:G170470; PIDN:AAA34177.1; PID:G170471

A:Experimental source: fruit

R:Sheehy, R.B.; Pearson, J.; Brady, C.J.; Hiatt, W.R.  
Mol. Gen. Genet. 208, 30-36, 1987

A>Title: Molecular characterization of tomato fruit polygalacturonase.  
A:Reference number: S06340

A:Accession: S06340

A:Molecule type: mRNA

A:Residues: 1-457 <SHE>

A/Cross-references: EMBL:X05656; NID:G19297; PIDN:CAA29148.1; PID:G19298

A>Note: Part of this sequence, including the amino end of the mature protein, was confir

R:Rose, R.E.; Houck, C.M.; Monson, E.K.; Dejeus, C.E.; Sheehy, R.E.; Hiatt, W.R.  
Nucleic Acids Res. 16, 7191, 1988

A>Title: The nucleotide sequence of the 5' flanking region of a tomato polygalacturonase  
A:Reference number: S02101; MUID:88303350; PMID:3405769

A:Accession: S02101

A:Molecule type: DNA

A:Residues: 1-93 <ROS>

A/Cross-references: EMBL:X07410; NID:G19295; PIDN:CAA30308.1; PID:G19296

R:Bird, C.R.; Smith, C.J.S.; Ray, J.A.; Mourreau, P.; Bevan, M.W.; Bird, A.S.; Hughes, S  
Plant Mol. Biol. 11, 651-662, 1988

A>Title: The tomato polygalacturonase gene and ripening-specific expression in transgen

A:Reference number: S08504

A:Accession: S08504

A:Molecule type: DNA

A:Residues: 1-457 <BR>

A/Cross-references: EMBL:X14074; NID:G19305; PIDN:CAA32235.1; PID:G295813

C:Comment: The polygalacturonase, a single cell wall enzyme, is the major enzyme medi

C:Superfamily: polygalacturonase

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-71/Domain: propeptide #status predicted <PRO>  
F:72-444/Product: polygalacturonase 2A #status experimental <MAT>

Query Match 100.0%; Score 2390; DB 2; Length 457;  
Best Local Similarity 100.0%; Pred. No. 9,6e-162;  
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MVIQRNSILLIIIFASISITCRSNVIDNLFKQVYNNILIEQFAHDFQAYLSLKNIE 60
DB 1 MVIQRNSILLIIIFASISITCRSNVIDNLFKQVYNNILIEQFAHDFQAYLSLKNIE 60
QY 61 SNNNIDKVDKNGIKVIVLSEFGAKGDKGTNDNIAFEQANNEACSSRPVQFVPPKNNYL 120
DB 61 SNNNIDKVDKNGIKVIVLSEFGAKGDKGTNDNIAFEQANNEACSSRPVQFVPPKNNYL 120
QY 121 LKQITFSGPCRSSISVYKIFGSLSEASSKISDYKDRRLMIAFDSVQNLVVGSGGTINGNGV 180
DB 121 LKQITFSGPCRSSISVYKIFGSLSEASSKISDYKDRRLMIAFDSVQNLVVGSGGTINGNGV 180
QY 181 WWPSSCKINKSLPCRDAPFLATLFWNCKNLKVNNLKSNAAQOIHFKESCTNVVASNLMIN 240
DB 181 WWPSSCKINKSLPCRDAPFLATLFWNCKNLKVNNLKSNAAQOIHFKESCTNVVASNLMIN 240
QY 241 ASAKSPMTDGVHVSNTQYIQTSDTIIGTGDPCISIVSGSQNVQATNITCGPHGISIGSL 300
DB 241 ASAKSPMTDGVHVSNTQYIQTSDTIIGTGDPCISIVSGSQNVQATNITCGPHGISIGSL 300
QY 301 GSGNSEAYSVNTVNEAKIIIGAENGVRKKTWQSGSGGASNIKFLNVMQDVKYPITIIDQN 360
DB 301 GSGNSEAYSVNTVNEAKIIIGAENGVRKKTWQSGSGGASNIKFLNVMQDVKYPITIIDQN 360
QY 361 YCQRVRECIQOFSAVQVQKVVYENIKGTSATKVAIKRDCSTNPFCEGIIENINLVGESG 420
DB 361 YCQRVRECIQOFSAVQVQKVVYENIKGTSATKVAIKRDCSTNPFCEGIIENINLVGESG 420
QY 421 KPSEATCKNVHPNNAEHVTPHCTSLSESEDEALLVNY 457
DB 421 KPSEATCKNVHPNNAEHVTPHCTSLSESEDEALLVNY 457

```

## RESULT 2

S31195  
polygalacturonase (EC 3.2.1.15) - avocado  
C/Species: Persea americana (avocado)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C/Accession: S31195; S28072  
R/Dopico, B.; Lowe, A.L.; Wilson, I.D.; Merodio, C.; Grierson, D.  
Plant Mol. Biol. 21, 437-449, 1993  
A>Title: Cloning and characterization of avocado fruit mRNAs and their expression during  
A/Reference number: S31195; MUID:93184201; PMID:8095163  
A/Accession: S31195  
A/Molecule type: mRNA  
A/Residues: 1-462 <DOP>  
A/Cross-references: EMBL:X66426  
R/Dopico, B.  
submitted to the EMBL Data Library, May 1992  
A/Reference number: S28072

A/Accession: S28072  
A/Molecule type: mRNA  
A/Residues: 1-181, 'S', 183-462 <DO2>  
A/Cross-references: EMBL:X66426; NID:g22630; PIDN:CAA47055.1; PID:g22631  
C/Superfamily: polygalacturonase  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 47.8%; Score 1142; DB 1; Length 462;  
Best Local Similarity 55.3%; Pred. No. 2.6e-73;  
Matches 224; Conservative 67; Mismatches 106; Indels 8; Gaps 5;

```

QY 28 DNNLFKQVYNNILIEQFAHDFQAYLSLKNIESNNNIDKVDKNGIKVIVLSEFGAKG 87
DB 40 DQRAYPTVFGPIID-EFSSIMGPEPSTLS-LERFNVGSGPETSPDDTIDISVDGAGADG 96

```

QY 88 KTYDNIAFEQANNEACSSRPVQFVPPKNNYLLKQITFSGPCRSSISVKIFGSLSEASSK 147  
DB 97 -TDYTAFEQANNEACSSGSLV--IYENKNYLLKQITFSGPCSSDLRKITGTEASSD 153

QY 148 ISDY--KDRRLMIAFDSVQNLVVGSGGTINGNGQVWPPSSCKINKSLPCRDAPFLATLFWN 205  
DB 154 QSDVWGNHRRKRWIEFEDISNLTLEGGGTINGNGETWWDSSCKRKSIPCKSAPFLATLFRS 213

QY 206 CKNLKVNNLKSNAAQOIHFKESCTNVVASNLMINAKSPNPDGVHVSNTQYIQTSDTI 265  
DB 214 CKNLVSDLSIKOSQKMHLSFDKQDVLASNLTWAPAEHSPNDGHIHTGRIHVMASV 273

QY 266 IGTGDDCISIVSGSQNVQATNITCGPHGISIGSLGSGNSEAYSVNTVNEAKIIIGAENG 325  
DB 274 IGTGDDCISIESGSKVVIATNITCGPHGISIGSLGDRNEAHVSGVLVQGNLFDITNG 333

QY 326 VRIKTWQSGGASNIKFLNVMQDVKYPITIIDQNYCDRAVEPCIQPSAVQKVVYENI 385  
DB 334 IRIKTWQSGGASKNIFQVNIWNNVNTPIIIDQYCDSDKDPCEQESAVQSVVAYNNI 393

QY 386 KGTSAATKVAIKRDCSTNPFCEGIIENINLVGESGRKSEATCKVY 430  
DB 394 RGTSASEVAVKFDCKSPCQGYIVGNINLVGNGKETTWSCSNI 438

## RESULT 3

T08215  
polygalacturonase (EC 3.2.1.15) 3 precursor - muskmelon

C/Species: Cucumis melo (muskmelon)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-May-2000  
C/Accession: T08215  
R/Hadfield, K.A.; Rose, J.K.; Yaver, D.S.; Berka, R.M.; Bennett, A.B.  
Plant Physiol. 117, 363-373, 1998

A>Title: Polygalacturonase gene expression in ripe melon fruit supports a role for polyg  
A/Reference number: Z16403; MUID:98289082; PMID:9625669  
A/Accession: T08215  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-461 <HND>

A/Cross-references: EMBL:AF062467; NID:g3320461; PID:g3320462  
A/Genetics:  
C/Genetics:  
A/Gene: MPg3  
C/Superfamily: polygalacturonase

C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-461/Product: polygalacturonase #status predicted <MAT>

Query Match 47.2%; Score 1128; DB 2; Length 461;  
Best Local Similarity 50.2%; Pred. No. 2.5e-72;  
Matches 224; Conservative 79; Mismatches 129; Indels 14; Gaps 7;

```

QY 7 SILLIIIFASISITCRSNVID--DNLFKQVYNNILIEQFAHDFQAYLSLKNIESNNNI 65
DB 17 TFLFVVVNFDTSTCGSYPDVNDNLPSSGDKEXEYGOYYHSSSEBSMTRT--RL 74

```

QY 66 DKVDKNGI--KVIVNVSEFGAKGDKGTNDNIAFEQANNEACSSRPVQFVPPKNNYLLK 122  
DB 75 EKWVSSPLASPEITNVDDYDGMGDEP-DIEAFETWKDACSNNAL-FLVPCORVHLK 132

QY 123 QITFSGPCRSSISVYKIFGSLSEASSKISDY-KDRRLMIAFDSVQNLVVGSGGTINGNGQV 181  
DB 133 PTFSGPCNSPLFLQIIGTAVASPHISDYEDKRHHWILFQMLGLRVEGKGIINGNGRKW 192

QY 182 WPPSSCKINKSLPCRDAPFLATLFWNCKNLKVNNLKSNAAQOIHFKESCTNVVASNLMINA 241  
DB 193 WUNSCKNKTLPCKEAPLAVTFYQCTNLRVGLFRNAQQMLSFQRCNNKALNLMITYA 252

QY 242 SAKSPMTDGVHVSNTQYIQTSDTIIGTGDPCISIVSGSQNVQATNITCGPHGISIGSLG 301  
DB 253 PGNSPMTDGHVHTQGTIVIKNCLIMTGDPCISIVSGSKVVRAGITTCGPHGISIGSLG 312

QY 302 GSGNSEAYSVNTVNEAKIIIGAENGVRKKTWQSGGASNIKFLNVMQDVKYPITIIDQNY 361  
DB 302 GSGNSEAYSVNTVNEAKIIIGAENGVRKKTWQSGGASNIKFLNVMQDVKYPITIIDQNY 361

```
Db 313 AGKSEAENVNVVDPAKFSGTSGNVCRIKTWQGGKGAQNIIFQNIWMNVNPIIIMQNY 372
QY 362 CDREVEPCIQFSAVQVKNVYENIKGTSATKVAIKEDCSTNPFCEGIIEMENILV----- 416
Db 373 CDQKECTQOADAIVANSNMVYQINIRGTSSEVAVKRDCSKYPCQGIILLQDINLVHKGN 432
QY 417 GESGKPESEATCKNVHFNNAEHVTPHC 442
Db 433 DDKSAQAEASCKNVKWKNGRVSPOC 458

RESULT 4
T46187
polygalacturonase (EC 3.2.1.15) precursor [similarity] - Arabidopsis thaliana
N:Alternate names: protein T8H10.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Aug-2000
C:Accession: T46187; T50674
R:Benes, V.; Rechmann, S.; Bokkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Accession: T46187
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <BEN>
A:Cross-references: EMBL:AF037367; PIDN:ACB6108.1
A:Experimental source: cultivar Columbia; BAC clone T8H10
R:Jenkins, E.S.; Roberts, J.A.
submitted to the EMBL Data Library, December 1997
A:Description: Deniscence-related expression of an Arabidopsis thaliana gene encoding a
A:Reference number: Z25172
A:Accession: T50674
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-431 <BEN>
A:Cross-references: EMBL:AF037367; PIDN:ACB6108.1
A:Experimental source: cultivar Landsberg erecta
C:Genetics:
A:Gene: T8H10.110
A:Map position: 3
A:Introns: 86/3; 129/3; 186/3; 193/3; 263/1; 290/2; 326/3; 364/3
C:Superfamily: polygalacturonase
C:Keywords: glycosidase; hydrolase
F:1-27/Domains: signal sequence #status predicted <Sig>
F:28-431/Product: polygalacturonase #status predicted <MAT>

Query Match 47.2%; Score 1127; DB 2; Length 431;
Best Local Similarity 49.6%; Pred. No. 2,7e-72;
Matches 223; Conservative 71; Mismatches 116; Indels 40; Gaps 6;

QY 7 SILLIIIFASSISTCRSNVID-----DNIEFKQVYDNILSOFPAHDFQAYLSYL 55
Db 9 AYVLCVLMISLCKALSSNVDDGFGHDSFESDSLKANNNDVL-----SLISSD 59
QY 56 SKNIESNNNIDKVDKNGIKVINVLSFGAKDGGTYNDINAEQAMNACSSRTPEVQFVVPK 115
Db 60 ETTLEAS-----TVSVSNFGAKDGGTDDTQAFKAMKACSTNGVTFLEVPK 107
QY 116 NKRYILLKQITFSPGCRSSISVVKIFGSLBASKISDYKDRMLWAFDSVQVLVVGGG--GT 173
Db 108 GKTYLLKSTFRFRSPCKSLRNFQILIGTLSASTKRSYDKKXHWLLEBVNLSIDGGSTGI 167
QY 174 INNGGVWMPSPCKINKSLPCRDAPALTATFWNCKNLKVNLLKSKNAQOIHKFESSCTNVV 233
Db 168 INNGKWTWONSCKIDKSKCTCAFTALITYNLKNANVKLRKYKMAQOIISIEKCNKVE 227
QY 234 ASNLMINASAKSPNTDGVHVSNTQYIQISDTIIGTDDCISIVSGSQNVQATNITCGPH 293
Db 228 VSNVEITAPDSDSNTDGIHTNTQNTIRVNSDIDGTGDDCISIDGTQNLIDFLTCGPH 287
QY 294 GISIGLSGNSLAVYNSVNVNNAKTIIGAENGVRIRIKTWQGGSGQASNIKFLNEMQDYK 353
Db 288 GISIGSLGDNSSKAYVSGINVDGAKFSESNDGVRIKTYQGGSGTAKNIKQNTIRENVKN 347
```

```
QY 354 PIIDQNYCDREVEPCIQFSAVQVKNVYENIKGTSATKVAIKEDCSTNPFCEGIIEMENI 413
Db 348 PIIDQDYCDK-DKEDQDSAVQVKNVYKNSIGTSATDVALTNLNCSEKPKCGIILENV 406
QY 414 NVGESGKPESEATCKNVHFNNAEHVTPHCT 443
Db 407 KIKG-----GTASCKNANVKNQGTVPKCS 431

RESULT 5
H84846
probable polygalacturonase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: H84846
R:Jin, X.; Kaul, S.; Rounsailey, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Wolfat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, S.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MID:20083487; PMID:10617197
A:Accession: H84846
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <STO>
A:Cross-references: GB:AE002093; NID:g2335094; PIDN:AA02763.1; GSPDB:GN00139
C:Genetics:
A:Gene: AE2941850
A:Map position: 2
C:Superfamily: polygalacturonase

Query Match 46.7%; Score 115.5; DB 2; Length 426;
Best Local Similarity 49.8%; Pred. No. 1,7e-71;
Matches 222; Conservative 74; Mismatches 117; Indels 33; Gaps 7;

QY 6 NSILLIIIFASSISTCRSNVIDNLFKQVY-----DNIEGFPAHDFQAYLSYLSKNIE 60
Db 6 NVTVFLMALMFWSCKASRISPNVYDHSYKFKSDSLIKRR--EDITGLRSFVNASLR 63
QY 61 SNNNIDKVDKNGIKVINVLSFGAKDGGTYNDINAEQAMNACSSRTPEVQFVVPKKNYL 120
Db 64 TPT-----TVSVSDFGAKDGGTDDTQAFNAMKACSSGAVNLVVKNTYL 112
QY 121 LKQITFSPGCRSSISVYKIFGSLSEASSKISDYKDRMLWAFDSVQVLVVGGG--GTINGNG 178
Db 113 LKSIGLTGCONSLITQIIGTLSASQKSDYKDISKIMFEDGVNLSVSGDGVVDGNG 172
QY 179 QVWMPSPCKINKSLPCRDAPALTATFWNCKNLKVNLLKSKNAQOIHKFESCTNVVASNLM 238
Db 173 EFWQNSCKRNKA-----KALTFFNSKSLIVKNLKVNAQOIISIEKSNVQVSNV 225
QY 239 INASAKSPNTDGVHVSNTQYIQISDTIIGTDDCISIVSGSQNVQATNITCGPHGISIG 298
Db 226 VTAPADSPMTDGIHTNTQNTIRVSESITGTGDDCISIESGQVQVINDITCGPHGISIG 285
QY 299 SLGSGSESEVNSVNVNNAEKTIIGAENGVRIRIKTWQGGSGQASNIKFLNEMQDYKYPID 358
Db 286 SLGDDNSKAFVSGVTVYDGAKLSTGDNVRIKTYQGGSGTASNLIFQNLQMDVKNPILID 345
QY 359 QNYCDREVEPCIQFSAVQVKNVYENIKGTSATKVAIKEDCSTNPFCEGIIEMENILVGE 418
Db 346 QDYCDK-SKCTTEKSAVQVKNVYRDISGSASENAITFCSKNYPCQGIIVLDRVNLKG- 403
QY 419 SGKPESEATCKNVHFNNAEHVTPHCTS 444
Db 404 ---GRACTNANVVKGAVALPQNS 425

RESULT 6
S71523
polygalacturonase (EC 3.2.1.15) [similarity] - peach
N:Alternate names: endopolygalacturonase
```

C/Species: Prunus persica (peach)  
 C/Date: 19-Mar-1997 #sequence revision 19-Mar-1997 #text\_change 31-Mar-2000  
 C/Accession: S71523; S71524; S71525  
 R/Lee, B.; Speirs, J.; Gray, J.; Brady, C.J.  
 A/Description: Homologies to the tomato endopolygalacturonase gene in the peach genome.  
 C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 21-Jul-2000  
 C/Accession: S71523  
 A/Reference number: S71523  
 A/Molecule type: DNA  
 A/Residues: 1-458 <LEB>  
 A/Cross-references: EMBL:X77231; NID:G479087; PIDN:CAA54448.1; PID:G479088  
 A/Experimental source: cultivar Maravilla  
 A/Note: This is a revision to the sequence from reference S71524  
 A/Submitter, D.R.; Speirs, J.; Orr, G.; Brady, C.J.  
 A/Submitted to the EMBL Data Library, January 1994  
 A/Description: Homologies to the tomato endopolygalacturonase gene in the peach genome.  
 A/Reference number: S71524  
 A/Accession: S71524  
 A/Molecule type: DNA  
 A/Residues: 53-92, 'I', '94-149, 'T', '151-185, 'W', '187-196, 'THA', '200, 'ESLS', '206-207, 'TWT', '212-  
 A/Cross-references: EMBL:X77231  
 A/Experimental source: cultivar Maravilla  
 A/Note: This sequence has been revised in reference S71523  
 A/Submitter, D.R.; Speirs, J.; Orr, G.; Brady, C.J.  
 A/Plant Cell Env. 13, 513-521, 1990  
 A/Title: Homologies to the tomato endopolygalacturonase gene in the peach genome.  
 A/Reference number: S71525  
 A/Accession: S71525  
 A/Molecule type: DNA  
 A/Residues: 265-458 <LEW>  
 A/Cross-references: EMBL:X77231  
 A/Experimental source: cultivar Maravilla  
 C/Genetics:  
 A/Intons: 107/3; 150/3; 204/3; 213/3; 283/1; 310/2; 386/3  
 C/Superfamily: polygalacturonase  
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 46.4%; Score 1108; DB 2; Length 458;  
 Best Local Similarity 48.9%; Pred. No. 6.5e-71;  
 Matches 226; Conservative 82; Mismatches 126; Indels 28; Gaps 8;

QY 1 MVIQRNSILLIIIPASSISTCRSNVIDNLFKQVYDNILOEFAHDFQAYLSYLSK 56  
 DB 1 MALQKHILVLFYVVSFCAASCYSGSQEVSLSHFVDH--EKESGVNSRAHPSMNTIEG 58

QY 57 -----KNBSNNIDKV--DKKGIVIVNVLSFGAKGDKTYDNIAPFQANMEAC 103  
 DB 59 VKEMEFIKPRAQFLPSSSRKLERAGSKSSSVKTISSANFAGAKNGAD-DTRAFERAKRAAC 117

QY 104 SSRTPVQFVVPKKNKNYLLKQITFSGPCRSISVYKIFGSLSEASSKISDYKDRMLIAPDSV 163  
 DB 118 SSKNAIYLVVVPQ-KTIVLRPIEFSGPKSHIMQITIEASDRSYKQVTHMLIDNV 176

QY 164 QNIVLVGGGGTINNGVWPPSSCKINKSLPCRD-APTALTFWNCKNLKYNLKSNAQOI 222  
 DB 177 QSLIVVPGTINNGNMMWENSCKRKPQPCNEQAPTAVTFNCKNLLVVKNLKIQDAQOM 236

QY 223 HIKEESTNVVASSLMLNLSAKSPNTDGVHVSNTQYQISPTIIGTDDICISYSGSQNV 282  
 DB 237 HVRFQNKVNEASHLVTAPEDESPNDGIHTNKNITTISSSVIGTDDICISYSGSQRV 296

QY 283 QATNITCGPGHGISIGLSGNSSEAVSVNTVNEAKITGAENGVRITKWOGSGSOASNIK 342  
 DB 297 QATDITGPGHGISIGLSGEDNNDHVSQVFNNAKISGTSNGRITKWOGSGSSASANTIV 356

QY 343 FLAVEMQDVKKYPIIIDQNYCD-RVEPCIQFSAVQVKNVYENIKGTSATKVAIKPDCST 401  
 DB 357 FQVVENMDVTPNPIIIDQNYCDHKTKDCRQSAVQVKNVLYONIRGTSASTDAITFCSQ 416

QY 402 NFPCEGIIMENINLVGSGKPSSEATCKNVFNNAEHTPHT 443  
 DB 417 SVPCQGIIVLQNIQD-----QNAKCKNNVPAKGVASPRCS 453

RESULT 7  
 T17011  
 polygalacturonase (EC 3.2.1.15) - apple tree  
 N/Alternate names: poly [1,4-a-D-galacturonide] glycan hydrolase  
 C/Species: Malus domestica (apple tree)  
 C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 21-Jul-2000  
 C/Accession: T17011  
 R/Atkinson, R.G.  
 A/Plant Physiol. 105, 1437-1438, 1994  
 A/Title: A cDNA clone for endopolygalacturonase from apple.  
 A/Reference number: Z18649; MUID:95062722; PMID:7972500  
 A/Accession: T17011  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-460 <ATK>  
 A/Cross-references: EMBL:L27743; NID:G456091; PIDN:AAA74452.1; PID:G456092  
 A/Experimental source: strain Golden delicious; ripe fruit  
 C/Function:  
 A/Description: catalyzes hydrolysis of 1,4-alpha-D-galactosiduronic linkages in galactur  
 A/Pathway: polysaccharide degradation  
 C/Superfamily: polygalacturonase  
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 45.6%; Score 1091; DB 2; Length 460;  
 Best Local Similarity 49.3%; Pred. No. 1.1e-69;  
 Matches 231; Conservative 76; Mismatches 126; Indels 36; Gaps 12;

QY 1 MVIQRNSILLIIIPASSISTCRSNVIDNLFKQVYDNILOEFAHDFQAYLSYLSK 57  
 DB 1 MALKTQILMSFVVFVVSFTTSCSGSQFQVNALHVSVDHVDKESGYSNRAVPSYTD 59

QY 58 NIE-----SNNDKVDKNGI-----KVYNLSFGAKGDKTYDNIAP 95  
 DB 60 TTEELKYMELIRPQDLPSSSRKMTI-TGGIATSSAPAKTISVDQFGKNGAD-DTQAF 117

QY 96 EOANNEACSSRTPVQFVVPKKNKNYLLKQITFSGPCRSISVYKIFGSLSEASSKISDYKDR 155  
 DB 118 VKAKKACSSSGAMVYVVPQ-KATVLRPIEFSGPKSHIMQITIEASDRSYKQID 176

QY 156 LMAIFDSQNLVVGGGGTINNGVWPPSSCKINKSLPCRD-APTALTFWNCKNLKYNL 214  
 DB 177 HMLIFDVQNLVVGPGTINNGNIMWKNCKIKPQPCGTVAPTAFTVFNCKNLLVKNL 236

QY 215 KSKNAQOIHKESECTNVVASSLMLNLSAKSPNTDGVHVSNTQYQISPTIIGTDDICIS 274  
 DB 237 NIQAQOIHVLFGCNIVQASCLVTAPEDSPNDGIHTVNTQITTISSVIGTDDICIS 296

QY 275 IVSGSQNVQATNITCGPGHGISIGLSGNSSEAVSVNTVNEAKITGAENGVRITKWOG 334  
 DB 297 IVSSSQNVQATDITGPGHGISIGLSGEDSEHVSQVFNNAKISGTSNGRITKWK 356

QY 335 SGAQSNIKFLAVEMQDVKKYPIIIDQNYCD-RVEPCIQFSAVQVKNVYENIKGTSATKV 393  
 DB 357 SGAATNIVFQNVQNDVTPNPIIIDQNYCDHKTKDCQKQSAVQVKNVLYONIRGTSASGD 416

QY 394 AIKEDCSNFPCEGIIMENINLVGSGKPSSEATCKNVFNNAEHTPHT 442  
 DB 417 AITLNCQSQVPCQGIIVLQSIQVQ--QNGR---ABCNNVQPAKGVASPRC 460

RESULT 8  
 D96833  
 hypothetical protein F18B13.25 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C/Accession: D96833  
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Hughes, B.; Huizart, L.  
 A/Genes: N.F.; Hughes, B.; Huizart, L.  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani,



Db 107 VERGPCQPHLSFKVDGTVIAQPPAPRWKSKIMLQFAQLTDFNLMTGVIDGQGGQMMAG 166  
QY 185 SCK-INKSLPC--RDAPALTFFWNNCKNLKVNLSKNAQOIHKEESCTNVASLMTNA 241  
Db 167 QCKVNVSTVCNDRRPRPAIKIDYKSVYKELTLMNSPEFLVGECEGYKIQGLKKA 226  
QY 242 SAKSPNTDGVHVSNTQYIQTSDTIIGTDDCISIVSGSQNVQATNITGPGHGISISLG 301  
Db 227 PRDSPNTDGDIFASKRPHIEKCVIGTGDICIAIGTSSNITIKDLIGPGHGISISLG 286  
QY 302 SGNEEAYVSNVTWVEAKITIGENGVRITKTWGGSGQASNIKFLNVEQDVXPIIIDONY 361  
Db 287 RDNRAEYSHVHVNAKFIIDTQNGLRITKQGGSLASYIYENEMTNSNPILINOFY 346  
QY 362 CDRVEPCIQFSAVQVKNVYENIKGTSATKVAIKFDCSTNPFCEGIMENINLVGESGK 421  
Db 347 CTSASACQNGRSAVQIDQVYKNIHGTSAATAIQLMCSDSVPCIGTQLSNVSLKLTSGK 406  
QY 422 PSEATCKNVHNNAEHVTPHCTSL 445  
Db 407 PASCVDKNARGFYSGRILPTCKNL 430

## RESULT 11

A96609  
probable polygalacturonase F25P12.85 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: A96609  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huntz, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: A96609  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-434 <STO>  
A/Cross-references: GB:AE051173; NID:G9954742; PID:AA09093.1; GSPDB:GN00141  
C/Genetics:  
A/Gene: F25P12.85  
A/Map position: 1  
C/Superfamily: polygalacturonase

Query Match 32.9%; Score 787.5; DB 2; Length 434;  
Best Local Similarity 41.3%; Pred. No. 3,1e-48;  
Matches 170; Conservative 68; Mismatches 155; Indels 19; Gaps 8;  
QY 54 YLSKNITSNNNIDKVDKNGIKVINLSPGAKGDKTYDNIAPFQAMNEACSSRTP--VQF 111  
Db 26 YLSPSPAPNPAVYNDNDIAPVDFVTSFGAIGDCSTDTSAFKMAWDAACMSTGPKALL 85  
QY 112 VVPEKNKYLLKQITFSGPCRSISVKTIFGSLA-----SKSDVKDRRLMAIPSYQNV 167  
Db 86 LVPTTFCFLKPTTFENPCRTNVLQIDGFIIVSDGPRSPWSNY--QRWMMFPRVNGLS 143  
QY 168 VVGSGGTINGNGQVWPPSSCKINKSL-----PCRDAPALTFFWNNCKNLKVNLSKNAQ 220  
Db 144 IQSGGVINGNGQKMMNLPCPKHKLNGTGTGFC--DSPVALRFLQSSKVRIGINPNASA 202  
QY 221 QIHKEESCTNVASLMTNINASAKSPNTDGVHVSNTQYIQTSDTIIGTDDCISIVSGSQ 280  
Db 203 QFHVRFPNCSDDVVDVYIIKAPASSPMTDGIHENTHNVQIRNMSINSGDCISIGAGCF 262  
QY 281 NVQATNITGPGHGISISIGSGNSEAYVSNVTWVEAKITIGENGVRITKTWGGSGQASN 340  
Db 263 NVDIKNTGCPSHGISISIGSGVHNSQAVSNITVTNSTIWNDSGVRITKTWGGSGSVSR 322

QY 341 IKELNVEQDVXPIIIDONYCDRVEPCIQFSAVQVKNVYENIKGT-SATKVAIKFDC 399  
Db 323 IVFSNITLVNVPNNPIMIDQYTC-QTNMNCANQTSATVISDVIYANKITDRLSPPIHFGC 381  
QY 400 STNPFCEGIMENINLVGESGKPSF-ATCKNVHFNNAHVTPHCTSLISED 450  
Db 382 SDSVCTNLTLEVDLFPSSKQHLNPFQWNAYSGMKLTVPVYGLLAPPD 433

## RESULT 12

S48730  
Cry j II protein - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 28-May-1999  
C/Accession: S48730  
R/Nanda, M.; Kurose, M.; Torigoe, K.; Hino, K.; Taniguchi, Y.; Fukuda, S.; Usui, M.; Kur  
FEBS Lett. 353, 124-128, 1994  
A/Title: Molecular cloning of the second major allergen, Cry j II, from Japanese cedar p  
A/Reference number: S48730; MUID:95010777; PMID:7926035  
A/Accession: S48730  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-514 <NMA>  
A/Cross-references: GB:D37765; NID:G577695; PID:BA07021.1; PID:d1007598; PID:G577696

Query Match 32.9%; Score 787.5; DB 2; Length 514;  
Best Local Similarity 38.5%; Pred. No. 3.9e-48;  
Matches 173; Conservative 78; Mismatches 163; Indels 35; Gaps 9;

QY 8 ILIIIIIPASSISITCRSNVIDNLEFKQYVDNILEQEFNAHDFQAVLYSKNIESNNNDK 67  
Db 12 VAMQLIIMAA-----EDQSAQIMLDSIDIQ-----YL-----RSRSIRK 47  
QY 68 VD---KNGIKVINLSPGAKGDKTYDNIAPFQAMNEACSSRTPVQVVPKKNYLLKQI 124  
Db 48 VEHSHDAINIFENVEKYGAVDGDGHDCTEAFSTAWQAC-KKPSAMLVPPNKKFVAVNL 106  
QY 125 TFSGPCRSISVKTIFGSLAASKISDYKDRRLMAIPSYQNVLVVGGGTINGNGQVWPPS 184  
Db 107 FPNGPCQPHFTFKVDGIIIAAQNPASWKNRITWQFAKLTGFTLMGKVIDGQGGQMMAG 166  
QY 185 SCK-INKSLPC--RDAPALTFFWNNCKNLKVNLSKNAQOIHKEESCTNVASLMTNA 241  
Db 167 QCKVNVSTVCNDRRPRPAIKIDYKSVYKELTLMNSPEFLVGECEGYKIIIGISITTA 226  
QY 242 SAKSPNTDGVHVSNTQYIQTSDTIIGTDDCISIVSGSQNVQATNITGPGHGISISLG 301  
Db 227 PRDSPNTDGDIFASKRPHIEKCVIGTGDICIAIGTSSNITIKDLIGPGHGISISLG 286  
QY 302 SGNEEAYVSNVTWVEAKITIGENGVRITKTWGGSGQASNIKFLNVEQDVXPIIIDONY 361  
Db 287 RDNRAEYSHVHVNAKFIIDTQNGLRITKQGGSLASYIYENEMTNSNPILINOFY 346  
QY 362 CDRVEPCIQFSAVQVKNVYENIKGTSATKVAIKFDCSTNPFCEGIMENINLVGESGK 421  
Db 347 CTSASACQNGRSAVQIDQVYKNIHGTSAATAIQLMCSDSVPCIGTQLSNVSLKLTSGK 406  
QY 422 PSEATCKNVHNN--AEHVTPHCTSL 448  
Db 407 --IASCLNDNANGYFSGHVPACKNLSPS 433

## RESULT 13

JC2498  
second major allergen Cry j II precursor - Japanese cedar  
C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 16-Mar-1995 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000  
C/Accession: JC2498; PC3346; A60147  
R/Komiyama, N.; Sone, T.; Shimizu, K.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 201, 1021-1028, 1994  
A/Title: cDNA cloning and expression of Cry j II, the second major allergen of Japanese  
A/Reference number: JC2498; MUID:94271186; PMID:8002972

A:Accession: J02498  
 A:Molecule type: mRNA  
 A:Residues: 1-514 <KOM>  
 A:Cross-references: DDBJ:D29772; NID:g506857; PIDN:BA06172.1; PID:g506858  
 A:Accession: PC2346  
 A:Molecule type: protein  
 A:Residues: 52-61 <KO2>  
 R:Sakaguchi, M.; Inouye, S.; Tanai, M.; Ando, S.; Usui, M.; Matubasi, T.  
 Allergy 45, 309-312, 1990  
 A:Title: Identification of the second major allergen of Japanese cedar pollen.  
 A:Reference number: A60147; MUID:90342988; PMID:2382797

A:Accession: A60147  
 A:Molecule type: protein  
 A:Residues: 55-64 <SAX>  
 C:Keywords: glycoprotein; pollen  
 F1:54/Domain: signal sequence #status predicted <Sig>  
 F1:55-460/Product: second major allergen Cry j #status predicted <MAT>  
 F1:429,460/472/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 32.9%; Score 787; DB 2; Length 514;

Best Local Similarity 40.5%; Pred. No. 4,2e-48;

Matches 168; Conservative 73; Mismatches 160; Indels 14; Gaps 7;

```

QY 45 AHDPAVL---SYLSKNIBSNNDKYD---KNGIVYNVLSFGAGDQKTYDNIAPDQA 98
DB 22 AEDQSAQIMDSVEKYLKRSNLSLRVESHSDHAINIFVVEKYGAVDGDHDCTEAFSTA 81
QY 99 KNEACSRPTVQFVVPKKNVYLKQITFSGPCSSISVKEIFGSLBASKISDYKDRMT 158
DB 82 WQACACNPS-AMLLVSGSKFVNNLFPNGPCOPHPTFGDGLIAYVONPASKNNRM 140
QY 159 AEDSVQNLVVGGGGTNGNGQVWVPSSCK-INKSLPC--RDAPTALTFFNCKNLKNNLK 215
DB 141 QFAKLITGPTLMGKGVLDGGCKQWMAQCKVWNGRELCNDRRTAIKFPDSTGLIIIGLK 200
QY 216 SKNAQOIHFKPESCTNVVANSMTNINASAKSPNTGVHVSNTQYIQISDTITIGDDICSI 275
DB 201 IAMSPEFHVFNGCEVKIIGISITRAPDSPTDJDIFASKNFHLOKMTIGTDDCVAI 260
QY 276 VSSQVQVQNTITCGPHGISIGSLGSGNSEAVSVTVTEAKITGAENVRKIKTWOGGS 335
DB 261 GTSSNIVIEDLTCGPHGISISLRNRSRAVSVHVNGAFPIDTQNLRLKTKWOGGS 320
QY 336 GQASNIKFLNVEMQDVKPIIIDQNYCDREPCIOQFSAVQVKNVYENIKGTSATKVAI 395
DB 321 GNASHIIVENINENENILINQFYCTASACQONRSANQIDVYKNIKRGTSATPAAI 380
QY 396 KPDGCTNFPCEGIIMENINLVGSGKPSKSEATCKNVHFN--AEHVTPHCTSLTIS 448
DB 381 QKCSDSMPCKDKLSDISIKLTSK--IASCLINDNANGVFSGHVTPACKNLSPS 433

```

#### RESULT 14

J02366

Jun a 2 protein - mountain cedar

C:Species: Juniperus ashei (mountain cedar)

C>Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C:Accession: J02366; PC7093

R:Tokuyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H.

Biochem. Biophys. Res. Commun. 275, 195-202, 2000

A:Title: Purification, identification, and cDNA cloning of Jun a 2, the second major all

A:Reference number: J02366

A:Accession: J02366

A:Molecule type: mRNA

A:Residues: 1-507 <YOK>

A:Cross-references: GB:AJ040453

A:Accession: PC7093

A:Molecule type: protein

A:Residues: 55-63 <Y02>

C:Comment: This protein, a second major allergen of mountain cedar pollen, which is inv

C:Keywords: glycoprotein; pollen

Query Match 32.0%; Score 764; DB 2; Length 507;

Best Local Similarity 41.8%; Pred. No. 1.8e-46;

Matches 157; Conservative 66; Mismatches 145; Indels 8; Gaps 5;

```

QY 75 VTNVLSFGAKGDKTYDNIAPDQANNEACSSRTPVQFVVPKKNVYLKQITFSGPCRSSI 134
DB 59 VFNVEHYGAVGDGKIDSTDAFEKTMMAACNKLNAV-FLVPANKKVVNNLVFPGFCQPH 117
QY 135 SVKIFGSLBASKISDYKDRRLMIAFDSVQNLVVGSGGTINGQVWVPSSCK-INKSLP 193
DB 118 SFKVDGTLIAYPDPAKMKSKIMHMFARLTDNLMTGVIIDQGNRMWSQCKTINGRTV 177
QY 194 GED--APVALTWNCKNLKVNKLKSKNAQOIHKEPESCTNVVANSMTNINASAKSPNTDV 251
DB 178 CNDKGRPTAIKIDFSKSVTKELTLNPSDFHVFEGECGVAKIQGKIKAPRDSPTMDGI 237
QY 252 HYSNTQYIOISDTITIGTGDPCISYVSGSNVQATNITCGPHGISIGSGNSEAVYSN 311
DB 238 DIFASKRFEIEKCTTGTGDCVAVGSGSNITTKDLCQPHGMSIGSLAKGSRSEVSF 297
QY 312 VTVNEAKITGAENGVARIKTWQSGQASNIKFLNVEMQDVKYPPIIIDQNYCDREPCIOQ 371
DB 298 VHLDDAKFTDTQNGLRIKTWQSGSLASHIITYENVMIAENPILINQFYCTSAACXQ 357
QY 372 FSAVQVKNVYENIKGTSATKVAIKFDCSTNFPCEGIIMENINLVGSGKPSKSEATCKNV 431
DB 358 RSAVKIQDVTFFKIHGTSATTAALQIMCSDSVCSNKLKSLNFKLULSGK--VATCVNKN 415
QY 432 ENN--AEHVTPHCTSL 445
DB 416 ANGYTNPILNPSCKSL 431

```

#### RESULT 15

B86368

protein P28C11.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B86368

R:Theologis, A.; Becker, J.R.; Palm, C.U.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: B86368

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1161 <STO>

A:Cross-references: GB:AE005172; NID:g8778576; PIDN:AAF79584.1; GSPDB:GN00141

A:Gene: P28C11.9

A:Map position: 1

Query Match 31.4%; Score 749.5; DB 2; Length 1161;

Best Local Similarity 36.1%; Pred. No. 5.8e-45;

Matches 163; Conservative 73; Mismatches 133; Indels 83; Gaps 8;

```

QY 71 NGIKVINVSFGAKGDKTYDNIAPDQANNEACSSRTPVQFVVPKKNVYLKQITFSGPC 130
DB 708 SGNLVNVDTPFGAGDGVSDDTQAFVSAMSKASTKSV-FLVPEGRRIYVNAIKFPGPC 766
QY 131 RSSISV-----KIFGSLBASKIS--DYKDRLTW 157
DB 767 EQLIIIVQKLSVSGKQLDHNVIYKTVLYNSFTCLCLQIDGTIVAPEPNSMBSKFORIW 826
QY 158 IAFDSVQNLVVGSGGTINGQVWVPSSCKINSLPDRDAPALTFFNCKNLKNNLKSK 217

```

```

Db      827 LEFSLKGVVYQKGKVIDGSGSKWMAASCKKXKS-----NALLIBSSGVXVSGLLIQ 873
QY      218 NAQQDHIKFBECTVNVVNSNLMINASAKSPNDGVYHVNSTQYIQISDPIIGEGDDCISIVS 277
      880 NSQQDNIIFLIPASDSRVKSKWVSSGDSBPNDGHHITGSTVILLQDCKIIGDDCVCSIVN 939
QY      278 GSONVQATNITCGFHGHISIGSLSGNSAEVSNVYVNEAKIIGAENVRIKTV----- 331
      940 AASNIKMNINICGEGHIGISIGSLGKMDNTGIVTVQVLDALLRETTNGRLIKTYQVHKKK 999
Db      332 -----QGSQASNIKFLNVMOQVKPFIITIDQNYCYRVE 366
QY      1000 KASLFSKNFDLARKSTTILFHCFPSFEEQGSQGVQGIREFNVAMOVANFILLDQYCSPT 1059
QY      367 PCIQQFSAVQYKVNYYENIKGTSAITKYAIKPDCSITNFCBGIIMENINILVGSSEKPSHAT 426
      1060 TCOMQTSVAKISQIIMYRNITGTTSAAKAIKPCASDTPVCSHIIVNVYLBENDQGVYAY 1118
Db      427 CKNVHNNAEH-----VTPHCTSLIISDEAL 453
      1119 C-----NSABGFGVTHFMSADCLYSHDDKGL 1145

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2004, 17:02:02 ; Search time 15.6574 Seconds  
(without alignments)  
1519.798 Million cell updates/sec

Title: US-10-018-604-2

Perfect score: 2390

Sequence: 1 MWIQRSILIIIFASIS.....VTPHCTSLIISDEALVNY 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2390	100.0	457	1 PGLR_LYCES	P05117 lycopersico
2	1295.5	54.2	467	1 PGLR_ACTCH	P35336 actinidia c
3	1136	47.5	462	1 PGLR_PERRA	Q02096 persea amer
4	1091	45.6	460	1 PGLR_MALDO	P48978 malus domes
5	787.5	32.9	514	1 MPAA_CRYUA	P43212 cryptomeria
6	708.5	29.6	422	1 PGLI_ARATH	P49062 arabidopsis
7	687.5	28.8	393	1 PGLR_PPRPE	P48979 prunus pers
8	667	27.9	407	1 PGLR_GOSHI	Q39786 gossypium h
9	667	27.9	444	1 PGLI_ARATH	P49063 arabidopsis
10	666	27.9	407	1 PGLR_GOSBA	Q39766 gossypium b
11	650.5	27.2	410	1 PGLT_MAIZE	P35339 zea mays (m
12	630	26.4	362	1 PGLR_OENOR	P24548 oenothera o
13	627	26.4	410	1 PGLR_MAIZE	P26215 zea mays (m
14	626	26.2	410	1 PGLS_MAIZE	P35338 zea mays (m
15	589.5	24.7	421	1 PGLR_MEDSA	Q40312 medicago sa
16	582.5	24.4	396	1 PGLR_TOBAC	Q05967 nicotiana t
17	572.5	24.0	397	1 PGLR_BRANA	P35337 brassica na
18	346	14.5	446	1 PGLX_COCCA	Q00359 cochliobolu
19	313	13.1	353	1 PGLR_ASFOR	P35335 aspergillus
20	311	13.0	370	1 PGLI_PENOL	Q9Y834 penicillium
21	311	13.0	435	1 PGLI_ASPTU	Q00293 aspergillus
22	310	13.0	363	1 PGLA_ASPTU	P41749 aspergillus
23	306	12.8	363	1 PGLR_ASPTU	P49575 aspergillus
24	300	12.6	364	1 PGLR_COCCA	P26215 cochliobolu
25	300	12.6	367	1 PGLR_PENDI	Q9Y718 penicillium
26	296	12.4	361	1 PGLR_YEAST	P47180 saccharomyc
27	291.5	12.2	402	1 PGLI_ERWCA	P26509 erwincia car
28	287.5	12.0	402	1 PGLI_ERWCA	P18192 erwincia car
29	285.5	11.9	368	1 PGLI_ASPTU	P26213 aspergillus
30	285	11.9	362	1 PGLI_ASPTU	P19805 aspergillus
31	280	11.7	376	1 PGLR_PENGR	Q93883 penicillium
32	278	11.6	363	1 PGLI_COLIN	Q00446 colletotric
33	277	11.6	380	1 PGLI_SCLIS	Q12708 sclerotinia

34	275	11.5	366	1 PGLB_ASPTU	P41750 aspergillus
35	273.5	11.4	371	1 PGLR_PENNA	Q42824 penicillium
36	271.5	11.4	373	1 PGLR_GIBFU	Q07181 gibberella
37	265	11.1	362	1 PGLI_ASPTU	P26214 aspergillus
38	256	10.7	380	1 PGLI_PENOL	Q9Y833 penicillium
39	247	10.3	383	1 PGLI_ASPTU	Q12554 aspergillus
40	242.5	10.1	378	1 PGLR_PENEN	Q59825 penicillium
41	241	10.1	440	1 RHGA_ASPTU	Q00001 aspergillus
42	237.5	9.9	602	1 PEHX_ERWCH	P15922 erwincia chr
43	186.5	7.8	531	1 PGLI_RALSO	P58598 raietonia s
44	183	7.7	312	1 PGLR_AGRTU	P27644 agrobacteri
45	181	7.6	529	1 PGLI_RALSO	P20041 raietonia s

## ALIGNMENTS

RESULT 1  
PGLR\_LYCES STANDARD; PRT; 457 AA.  
ID P05117;  
AC 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Polygalacturonase 2A precursor (EC 3.2.1.15) (PG-2A) (pectinase).  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Ailsa Craig;  
RA Bird C.R., Smith C.J.S., Ray J.A., Mourreau P., Bevan M.W.,  
RT Bird A.S., Hughes S., Morris P.C., Grierson D., Schuch W.,  
RT "The tomato polygalacturonase gene and ripening-specific expression in  
transgenic plants.";  
RL Plant Mol. Biol. 11:651-662(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Ailsa Craig;  
RX MEDLINE=8706731; PubMed=3786135;  
RA Grierson D., Tucker G.A., Keen J., Ray J., Bird C.R., Schuch W.,  
RT "Sequencing and identification of a cDNA clone for tomato  
polygalacturonase.";  
RL Nucleic Acids Res. 14:8595-8603(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Sheehy R.E., Pearson J., Brady C.J., Hiatt W.R.,  
RT "Molecular characterization of tomato fruit polygalacturonase.";  
RL Mol. Gen. Genet. 208:30-36(1987).  
RN [4]  
RP REVISIONS.  
RA Hiatt W.R.,  
RT Submitted (OCT-1987) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Acts in concert with the pectinesterase, in the ripening  
process. Is involved in cell wall metabolism, specifically in  
polyuronide degradation.  
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
galactosiduronic linkages in pectate and other galacturonans.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- DEVELOPMENTAL STAGE: In ripening fruit.  
CC -!- BIOTECHNOLOGY: The effect of PG can be neutralized by introducing  
an antisense PG gene by genetic manipulation. The Flavr Savr  
tomato produced by Calgene (Monsanto) in such a manner has a  
longer shelf life due to delayed ripening.  
CC -!- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, X14074; CAA32235.1; -

DR EMBL, M37304; AAA34178.1; -

DR EMBL, X04583; CAA28254.1; -

DR EMBL, X05656; CAA29148.1; -

DR EMBL, A15981; CAA01256.1; -

DR EMBL, A24194; CAA01720.1; -

DR PIR, A25534; A25534.

DR InterPro; IPR000743; Glyco\_hydro\_28.

DR InterPro; IPR006626; PDI.

DR Pfam; PF00295; Glyco\_hydro\_28; 1.

DR SMART; SMO0710; Pdi1; 4.

DR PROSITE; PS00502; POLYGALACTURONASE; 1.

KW Hydrolyase, Glycosidase; Cell wall; Signal; Fruit ripening;

KW Glycoprotein; Genetically modified food.

FT SIGNAL 1 24 POTENTIAL.

FT PROPEP 25 71 POTENTIAL.

FT CHAIN 72 457 POLYGALACTURONASE 2A.

FT ACT SITE 293 293 PROBABLE. (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 457 AA; 50051 MW; 449E4DC36919B074 CRC64;

Query Match 100.0%; Score 2390; DB 1; Length 457;  
Best Local Similarity 100.0%; Pred. No. 1,6e-160;  
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIORNSILLIIIFASISITCRSNVIDDNLFKQYVDNLTBOEPAHDPQAYLSLTKITE 60  
Db 1 MVIORNSILLIIIFASISITCRSNVIDDNLFKQYVDNLTBOEPAHDPQAYLSLTKITE 60

QY 61 SNNNIDKVDKNGIKVINLVSFGAKDGKTYDNIAFEQAMNACSSRTVPQFVPEKKNYLT 120  
Db 61 SNNNIDKVDKNGIKVINLVSFGAKDGKTYDNIAFEQAMNACSSRTVPQFVPEKKNYLT 120

QY 121 LKQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAFPSYONLVVGGGGTNGNGQV 180  
Db 121 LKQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAFPSYONLVVGGGGTNGNGQV 180

QY 121 LKQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAFPSYONLVVGGGGTNGNGQV 180  
Db 121 LKQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAFPSYONLVVGGGGTNGNGQV 180

QY 181 WMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQOIHIFKFSCTVVASNLMIN 240  
Db 181 WMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQOIHIFKFSCTVVASNLMIN 240

QY 241 ASAKSPNTDGVHVSNTQYIQISDTIIGTGDCCISIVSGSQNVQATNTTCGFGHGISIGSL 300  
Db 241 ASAKSPNTDGVHVSNTQYIQISDTIIGTGDCCISIVSGSQNVQATNTTCGFGHGISIGSL 300

QY 301 GSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSNKFLNVEMQDYKTYPIIIDON 360  
Db 301 GSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSNKFLNVEMQDYKTYPIIIDON 360

QY 361 YCDREVEPCIOQFSAVQKYNVYENIKGTSATKVAIKFDCSTNFCDEGIMENIMLVGESSG 420  
Db 361 YCDREVEPCIOQFSAVQKYNVYENIKGTSATKVAIKFDCSTNFCDEGIMENIMLVGESSG 420

QY 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNNY 457  
Db 421 KPSEATCKNVHFNNAEHVTPHCTSLSEDEALLNNY 457

RESULT 2  
PGLR ACTCH STANDARD; PRT; 467 AA.

AC P35336;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Polyalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase).  
OS *Actinidia chinensis* (Kiwi) (Yangtze).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
CC Ericales; Actinidiaceae; Actinidia.  
CC NCBI\_TaxID=3625;  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=cv. Deliciosa;  
CC MEDLIN=94302157; PubMed=8029342;  
CC Atkinson R.G., Gardner R.C.;  
CC "A polyalacturonase gene from kiwifruit (*Actinidia deliciosa*).";  
CC Plant Physiol. 103:669-670(1993).  
CC -1- FUNCTION: Acts in concert with the pectinesterase, in the ripening  
CC process. Is involved in cell wall metabolism, specifically in  
CC polyuronide degradation.  
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
CC galactosiduronic linkages in pectate and other galacturonans.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DEVELOPMENTAL STAGE: In ripening fruit.  
CC -1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, L12019; AAC14453.1; -

DR InterPro; IPR000743; Glyco\_hydro\_28.

DR Pfam; PF00295; Glyco\_hydro\_28; 1.

DR PROSITE; PS00502; POLYGALACTURONASE; 1.

KW Hydrolyase, Glycosidase; Cell wall; Signal; Fruit ripening;

KW Glycoprotein.

FT SIGNAL 1 27 POTENTIAL.

FT CHAIN 28 467 POLYGALACTURONASE.

FT ACT SITE 306 306 PROBABLE.

FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 467 AA; 50776 MW; 5A9A61483C028B7A CRC64;

Query Match 54.2%; Score 1295.5; DB 1; Length 467;  
Best Local Similarity 54.6%; Pred. No. 1.1e-83;  
Matches 257; Conservative 75; Mismatches 112; Indels 27; Gaps 7;

QY 1 MVIORNSILLIIIFASISITCRSNVIDDNLFKQYVDNLTBOEPAHDPQAYLSLTKITE 58  
Db 1 MVIORNSILLIIIFASISITCRSNVIDDNLFKQYVDNLTBOEPAHDPQAYLSLTKITE 58

QY 59 IESNNNIDKVDKNGI-----KVINLVSFGAKDGKTYDNIAFEQAMN 101  
Db 59 IESNNNIDKVDKNGI-----KVINLVSFGAKDGKTYDNIAFEQAMN 101

QY 102 ACSSRTVPQFVPEKKNYLTQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAF 160  
Db 102 ACSSRTVPQFVPEKKNYLTQITFGPGCRSSISVYIFGSLSEASKISDYKDRRLWAF 160

QY 115 ACSSTSAVLLVPEK-KVYLVPISFSGCKSGLTQITGRTIASDDSDYKDRHMLVF 173  
Db 115 ACSSTSAVLLVPEK-KVYLVPISFSGCKSGLTQITGRTIASDDSDYKDRHMLVF 173

QY 161 DSYONLVVGGGGTNGNGQVWMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQ 220  
Db 161 DSYONLVVGGGGTNGNGQVWMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQ 220

QY 174 DSYONLVVGGGGTNGNGQVWMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQ 223  
Db 174 DSYONLVVGGGGTNGNGQVWMPSSCKINKSLPRDAPALTLPNNCKNLKYNLKSXAQ 223

QY 221 QIHIFPESCTNVVASNLMINASAKSPNTDGVHVSNTQYIQISDTIIGTGDCCISIVSGSQ 280  
Db 221 QIHIFPESCTNVVASNLMINASAKSPNTDGVHVSNTQYIQISDTIIGTGDCCISIVSGSQ 280

QY 234 QIHVSPNCNVVQASNLMTAPENSPTMDGIHYGTQNIHSSCVIIGTGDCCISIVGSR 293  
Db 234 QIHVSPNCNVVQASNLMTAPENSPTMDGIHYGTQNIHSSCVIIGTGDCCISIVGSR 293

QY 281 NVQATNTTCGFGHGISIGSLGSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSN 340  
Db 281 NVQATNTTCGFGHGISIGSLGSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSN 340

QY 294 KVRVNDITCGFGHGISIGSLGSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSN 353  
Db 294 KVRVNDITCGFGHGISIGSLGSGNSEAVYVNTVNEAKITGAENGVRKTMQSGGSAQSN 353

QY 341 IKFLNVEMQDYKTYPIIIDONVCDREVEPCIOQFSAVQKYNVYENIKGTSATKVAIKFDCS 400  
Db 341 IKFLNVEMQDYKTYPIIIDONVCDREVEPCIOQFSAVQKYNVYENIKGTSATKVAIKFDCS 400

QY 354 IKFQNVEMHVENITPIIIDONVCDODKPCQEGSSAVQKYNVYENIKGTSATKVAIKFDCS 413  
Db 354 IKFQNVEMHVENITPIIIDONVCDODKPCQEGSSAVQKYNVYENIKGTSATKVAIKFDCS 413

QY 401 TNFCDEGIMENIMLVGESSGKPSEATCKNVHFNNAEHVTPHCTSLSEDE 451

Dd		414	KRFPCQGVLEVDVLEIFGGAANAANKNLNVELSETGVSPHCQEGBGEER	464
RESULT 3				
PGLR_PERAE	ID	PGLR_PERAE	STANDARD;	PRT; 462 AA.
AC	Q02096;			
DT	01-JUL-1993	(Rel. 26, Created)		
DT	01-JUL-1993	(Rel. 26, Last sequence update)		
DE	10-OCT-2003	(Rel. 42, Last annotation update)		
DR	Polygalacturonase precursor (EC 3.2.1.15) (PG)	(Pectinase).		
OS	Pearsea americana (Avocado).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; magnoliids; Laurales; Lauraceae;			
OX	NCHI_TaxID=3435;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	STRAIN=cv. Hass; TISSUE=Pericarp;			
RX	MEDLINE=93184201; PubMed=8095163;			
RA	Dopico B., Lowe A.L., Wilson I.D., Merodio C., Grierson D.;			
RT	"Cloning and characterization of avocado fruit mRNAs and their			
RL	expression during ripening and low-temperature storage.";			
PL	Plant Mol. Biol. 21:437-449 (1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Hass; TISSUE=Mesocarp;			
RX	MEDLINE=94269193; PubMed=8208850;			
RA	Katsunai S.Y., Lin A.C., Percival F.W., Lacies G.G.,			
RT	Christoffersen R.E.;			
RL	"Ripening-related polygalacturonase cDNA from avocado.";			
CC	-!- FUNCTION: Acts in concert with the pectinesterase, in the ripening			
CC	process. Is involved in cell wall metabolism, specifically in			
CC	polyuronide degradation.			
CC	-!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-			
CC	galactosiduronic linkages in pectate and other galacturonans.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- DEVELOPMENTAL STAGE: In ripening fruit.			
CC	-!- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.			
CC	-----			
CC	This SWISS-PROT entry is copy-right. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license@isb-stb.ch)			
CC	or send an email to license@isb-stb.ch)			
CC	-----			
EMBL:	X66426; CAA47055.1; --			
EMBL:	L06094; AAA32914.1; --			
PIR:	G31195; SA1195.			
IrcPro:	IPR00743; Glyco_hydro_28.			
IrcPro:	IPR00626; Pbh1.			
Pfam:	PF00295; Glyco_hydro_28; 1.			
SMART:	SM00710; Pbh1; 5.			
PROSITE:	PS00502; POLYGALACTURONASE; 1.			
Hydrolase:	Glycosidase; Cell wall; Signal; Fruit ripening;			
KM	Glycoprotein.			
FT	SIGNAL	1	22	POTENTIAL.
FT	CHAIN	23	462	POLYGALACTURONASE.
FT	ACT SITE	301	301	BY SIMILARITY.
FT	CARBOND	294	294	N-LINKED (GLCNAC... ) (POTENTIAL).
FT	CONFLICT	182	182	S->I (IN REF. 2).
FT	CONFLICT	416	419	YIVG->ILHE (IN REF. 2).
FT	CONFLICT	442	462	LIREGLSTFLFMKRVHSCSY->TTGGKYPPSCL
SEQ	SEQUENCE	462 AA;	50290 MW;	ABCF068ACCB17A8 CRC64;
Query Match	Best Local Similarity	47.5%;	Score 1136;	DB 1; Length 462;
Matches 223;	Conservative	55.1%;	Pred.No.1.7e-72;	Indels 8; Gaps 5,

Qy	28	DNLFKQYKDNLLEOEFDHDEQATLSTYSKNIENSNNDKYDKMGIKVYNLVSFGAKDG	87
Db	40	DORAPYPIFGILD-EFBSIMGFEBSILS--LERNPVGPEPITSPTDLSVDFGARGDG	96
Qy	88	KTYNDIAFEQAMNACSSRTPEVQFVPPKRYKYLKQITFSGCRSSISVKIFGSLASAK	147
Qy	148	ISDY--KDRLMIAPDSYQNLVVGSGGCTINGNGQVWVWSSCKINSLPERDAPTLTFMN	205
Db	154	QSDWVGHRKRWIIEFEDISNLTLLGGGTSNGGEMWSSCKRKSLLPCKSPATLTPRS	213
Qy	206	CKNLKNNVLKSKNQOIHIFESCTNNVASNLMIAASAKSPYTDGVHVSNTQYIQTSDPI	265
Db	214	CKNLIVSPLSLKSDQKHLSPDKCDVIALSNLMVAPRHPSPITDGIHITGRIHYMNSV	273
Qy	266	IGTGDDCISIVSGSQWQVATNITCGPGHGISTGSGNSGSEAYSNVTVNKAIIIGAENG	325
Db	274	IGTGDDCISIESGSRMVAIVATNITCGPGHGISTGSLGDRNSEAHVSGVLDGGLPDTNG	333
Qy	326	VRITWGGSGSQASNIKFLANVMQDVKPIITIDQYQGRVPECIOGFSAVQKVVYXNT	385
Db	334	LRIWGGSGSSAKNIKQNTVMHVNTPITIDYQCSKDPCEQBSXAKVSNVAYMNI	393
Qy	386	KGSATKVAIKFDCSTNPFCEGIIMENINLVGSSCKPEATCKNV	430
Db	394	RGTSAEVAIVAFDCSKSPCGYIVGNINIVNGSGKETTMSCSNI	438

RESULT 4

PELR	MALDO	STANDARD	PRT	460 AA.
AC	P48978;			
DT	01-FEB-1996	(Rel. 33, Created)		
DT	01-FEB-1996	(Rel. 33, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Polygalacturonase precursor (EC 3.2.1.15) (PG (Pectinase).			
OS	Malus domestica (Apple) (Malus sylvestris).			
OC	Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;			
OC	Spermatophyta, Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid1; Rosales; Rosaceae; Maloideae; Malus.			
OX	NCBI_TaxID=3750;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=cv. Golden Delicious;			
EX	MEDLINE=95062722; PubMed=7972500;			
RA	Atkinson R.G.;			
RT	"A cDNA clone for endopolygalacturonase from apple.";			
RL	Plant Physiol. 105:1437-1438(1994).			
CC	-!- FUNCTION: Acts in concert with the pectinesterase, in the ripening			
CC	process. Is involved in cell wall metabolism, specifically in			
CC	polyuronide degradation.			
CC	-!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-			
CC	galactosiduronic linkages in pectate and other galacturonans.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboratio			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercia			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	PIR, I27743; AAA74452.1; -.			
DR	PIR, I17011; T17011.			
DR	InterPro: IPR000743; Glyco hydro 28.			
DR	Pfam: PF00295; Glyco hydro 28; 1.			
DR	PROSITE: PS00502; POLYGALACTURONASE; 1.			
KM	Hydrolase; Glycosidase; Cell wall; Signal; Fruit ripening;			
KM	Glycoprotein			

```

FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 460 POLYGALACTURONASE.
FT ACT SITE 315 315 PROBABLE. (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED. (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED. (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 460 AA: 49403 MW; 13E99A7A049FAFCE CRC64;

Query Match 45.6%; Score 1091; DB 1; Length 460;
Best Local Similarity 49.3%; Pred. No. 2.4e-69;
Matches 231; Conservative 76; Mismatches 126; Indels 36; Gaps 12;

QY 1 MWIGNSILLIIFASSIST--CRSNVTD-NLFKQYVDNLEQFAHDFQAYLSYSK 57
D 1 MALKTQLMSFVVFVVFSTSCSGSFGQVNMALSHVDHVDKESGYNRAVSY-TD 59
QY 58 NIE-----SNNNIDKVDKNGI-----KVINLVSPFAGKGGKTYDIAF 95
D 60 TIEGLKWEILIRPTQLFSSRKLMTI-TGGIATSSAPAKTISVDDFGKKGAD-DTQAF 117
QY 96 EQANNEACSSRTPOFVVPKKNKYLKQITFGPCRSISVTKIFGSLASXISDYKDR 155
D 118 VKAKKACSSSGANLVLPQ-KNVLVRIEFGSPCKSQLTIQITGTEASEDRSTYKDID 176
QY 156 LMIAPDSVQNLVGGGTTINGQVWPPSSCKINKSLPCRD-APTALTFWCKNLKXNLT 214
D 177 HMLIFDVQNLVVGPTINGNINMWKNSCKIKPQPCGTAPPAVTFRCNNLVVKNL 236
QY 215 KSKKAOQIHIFESCTVVAANLMMASAKSPNTDGVVSTQYIQTSDITIGGDCIS 274
D 237 NIQAQGHVIFQNCINQASCLTVLAPEDSPNDGHTVNTQNTITSSVIGTDCIS 296
QY 275 IVSSGQVQATNITCGFGHGISIGSLGSEAVSVNTVEAKTIGANGVRIKTWQG 334
D 297 IVSSGQVQATNITCGFGHGISIGSLGSEAVSVNTVEAKTIGANGVRIKTWQG 356
QY 335 SGQASNIKFLVENQDYKPIIIDQNYCD-RVEPCIQGFSAYQVKNVYENIKTSATKV 393
D 357 SGSAKTNIIVQVQNDVNTPIIIDQNYCDHKTQCKQKSAVQKAVLVQNIKRTGSASGD 416
QY 394 AIKFDGCTNPFCEGIMENINLVGSEKPSATCKNHNNAHEVTPHC 442
D 417 AITLNGSQSVPCGIVLQSVOL--QNGR--AECNNVQPAYKGVVSPRC 460

RESULT 5
MPA2_CRYTA STANDARD; PRT; 514 AA.
AC PA3212;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Possible polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase)
DE (Major pollen allergen Cry j 2) (Cry j II).
OS Cryptomeria japonica (Japanese cedar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Cryptomeria.
OX NCBI_TaxID=3369;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Pollen;
RX MEDLINE=95010777; PubMed=7926035;
RA Namba M., Kurose M., Torigoe K., Hino K., Taniguchi Y., Fukuda S.,
RA Usui M., Kurimoto M.;
RT "Molecular cloning of the second major allergen, Cry j II, from
RL Japanese cedar pollen.";
RL FEBS Lett. 353:124-128(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RX MEDLINE=94271186; PubMed=8002972;
RA Komiyaama N., Sone T., Shimizu K., Morikubo K., Kano K.;
RT "cDNA cloning and expression of Cry j II the second major allergen of
RL Japanese cedar pollen.";
```

```

RU Biochem. Biophys. Res. Commun. 201:1021-1028(1994).
RN [3]
RP SEQUENCE OF 55-64.
RX MEDLINE=90342988; PubMed=2382797;
RA Sakaguchi M., Inoue S., Tanai M., Ando S., Usui M., Matubasi T.;
RT "Identification of the second major allergen of Japanese cedar
RL pollen.";
RL Allergy 45:309-312(1990).
CC -I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
CC galactosiduronic linkages in pectate and other galacturonans.
CC -I- SUBCELLULAR LOCATION: SECRETED OR AMYLOPLAST (POTENTIAL).
CC -I- ALLERGEN: Causes an allergic reaction in human.
CC -I- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; D37765; BAA07021.1; -
DR EMBL; D29772; BAA06172.1; -
DR PIR; JC2498; JC2498.
DR PIR; S48730; S48730.
DR HSSP; P26509; 1BHE.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR006626; PDI.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR SMART; SM00710; PdiL; 5.
DR PROSITE; PS00502; POLYGALACTURONASE; 1.
KW Hydrolyase; Glycosidase; Cell wall; Signal; Zymogen; Fruit ripening;
KW Amyloplast; Glycoprotein; Allergen.
FT SIGNAL 1 22
FT PROPEP 23 45
FT CHAIN 46 433
FT PROPEP 434 514
FT ACT SITE 278 278
FT CARBOHYD 460 460
FT CARBOHYD 472 472
FT CONFLICT 5 5
FT CONFLICT 12 12
FT CONFLICT 34 35
FT CONFLICT 37 37
FT CONFLICT 88 88
FT CONFLICT 98 98
FT CONFLICT 451 451
FT CONFLICT 454 454
FT CONFLICT 504 504
FT CONFLICT 507 507
SQ SEQUENCE 514 AA: 56645 MW; 624611C3FA8D6302 CRC64;

Query Match 32.9%; Score 787.5; DB 1; Length 514;
Best Local Similarity 38.5%; Pred. No. 5.5e-48;
Matches 173; Conservative 78; Mismatches 163; Indels 35; Gaps 9;

QY 8 ILLIIFASSISTCRSNVTDNLFKQYVDNLEQFAHDFQAYLSYSKRIENNNIDK 67
D 12 VAWQLITMAA-----EDQSAQIMLDSIDIO-----YL-----RNRSLRK 47
QY 68 VD---KNGIKYINLVSGAKGDKTYDNIAPQANNEACSSRTPOFVVPKKNKYLKQI 124
D 48 VEHSHRHAINIFNVEKYGAVDGDHCTEAFSTMQAAC-KPSAMLVIPANKGFVYVNL 106
QY 125 TFGSPCRSSISVKIFGSLASXISDYKDRRLMIAPDSVQNLVVGSGGTTINGQVWPPS 184
D 107 FFGNPPCQPHFTFKVDGIIAIVQNPASWKNRIRIWFQAKLTGFTLMGKGVIDGQKQWAG 166
QY 185 SKC-INKSLPC--PDAPTALTFWCKNLKXNLTSKKAOQIHIFESCTVVAANLMMNA 241
D 167 QCKWVNGREICNDRDPTAIKFDSTGLIITGLKLMNSPEHVLVFGNCEGVKIIIGISIT 226
```

```
QY 242 SAKSPNTDGVHVNIOYIQTISTITGDDCISIVSGSNVQATNITCGPBGHSIGSISG 301
DB 227 PRDSEPTDIDIFASKRPHLQKNTIGTDDCVAIGTSSNIVIEDICGPHGISIGSISG 286
QY 302 SGNSEAVSNVYVNEAKITGAENGVRKTMGGSGGASNKIKFLNVMQVKKYITIDONY 361
DB 287 RENSREAVSVYHVGAKFIQTQNGRIKTMGGSGGASHITVENEWIMENSENITLINOQF 346
QY 362 CDREVEPCIQFSAVQVKNVYENIKGTSATKVAIKFDCSTNPFCEGILIMENITLVGSEK 421
DB 347 CTSASACQORSAVQVQDVTYKRIKRTSATMAAIOUKCDSMCKDKIKSLKTSK 406
QY 422 PSEATCKNVHFN--AEHVTPHCTSLSEIS 448
DB 407 --IASCLINDNANGYFSGHVTIPACKNLSPS 433

RESULT 6
PG1L ARATH STANDARD; PRT; 422 AA.
ID PG1L ARATH
AC P49062;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exopolylacturonase clone GBGE184 precursor (EC 3.2.1.67) (ExoPG)
DE (pectinase) (Galacturan 1,4-alpha-galacturonidase).
GN PG43 OR AT1G02790 OR T14P4.2 OR F22D16.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. C24; TISSUE=flower buds;
RA MEDLINE=99413295; PubMed=10465285;
RX Torik M., Mandaron P., Thomas F., Quigley F., Mache R., Falconet D.;
RT "Differential expression of a polygalacturonase gene family in
RT Arabidopsis thaliana";
RL Mol. Genet. 261:948-952(1999).
RN
RP SEQUENCE FROM N.A.
RC Torik M., Thomas F., Mache R., Mandaron P., Falconet D.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA MEDLINE=21016719; PubMed=11130712;
RX Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Brehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Felblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hitzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetschka I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Ienz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzalli A.,
RA Millischer J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
RA Pai G., Peterson U., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.U., Town C.D.,
RA Utenback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
CC -1- FUNCTION: May function in depolymerizing pectin during pollen
CC development, germination, and tube growth. Acts as an exo-
CC polygalacturonase.
CC -1- CATALYTIC ACTIVITY: (1,4)-alpha-D-galacturonide (N) + H(2)O =
CC (1,4)-alpha-D-galacturonide (N-1) + D-galacturonate.
CC -1- SUBCELLULAR LOCATION: Secreted.
```

```
CC -1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; X72291; CA451032.1; -.
DR EMBL; Y16230; CA476127.1; -.
DR EMBL; AC009525; AA02888.1; -.
DR PIR; S34199; S34199.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR006626; Pbh1.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR SMART; SM00710; Pbh1; 5.
DR PROSITE; PS00502; POLYGALECTURONASE; 1.
DR HydroLase; Glycosidase; Cell wall; Signal; Glycoprotein;
KM Multigene family.
FT STGNAL 1 31 POTENTIAL.
FT CHAIN 32 422 EXOPOLYLACTURONASE CLONE GBGE184.
FT ACT_SITE 264 264 PROBABLE.
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 422 AA; 44430 MW; 17B02E2CE979BA6A CRC64;

Query Match 29.6%; Score 708.5; DR 1; Length 422;
Best Local Similarity 40.8%; Pred. No. 1,5e-42;
Matches 151; Conservative 66; Mismatches 128; Indels 25; Gaps 11;

QY 75 VINVLSFGAKGDKXTDNIAFEQAMNEACSSRPVQVFPKRNKNTLYLKQITFSGPCRSI 134
DB 51 VYDITKFGAVGDSSTTFKAFLLNTWVQVCDSPVPALILVPKG-TPLAGPITFAPGPKSKV 109
QY 135 SVKIFGSLSEASSKISDPKRRMLIAPDSYQNVVGGGGITNGQ-VWMPSSCKINKSLP 193
DB 110 TVNVIGTIIAT--SGVATPE-WFLPERVDNLVLTGTFPHGKEAWKADGC--GKVVQ 164
QY 194 CRDAPFALTFPKNCKNLKVNLIKSKNAQQIHKESCTNNVASULMINKAKSPRTGSVHV 253
DB 165 CNLPPTSLKFRNKNKEVINGISSVNAKAFMFLVKTENNVIQIKLTAPAESPTDSIHL 224
QY 254 SNTQYIQTSTITGDDCISIVSGSNVQATNITCGPBGHSIGSISGSNSZAYVSNVT 313
DB 225 SNADNVSLDSTIATGDDCVSVGRSSNVTVERVITGPHGLSVGSLGKRYNEEDVSGIH 284
QY 314 VNEAKITGAENGVRKIKTMGGSG--GQASNIKFLNVMQVKKYPIIIDONYCDREVEPCIQ 371
DB 285 VNNCTMIENLNDNGRIKTM--GGSDPSKAVDIKFENITLMQSVKNPIIIDONYSGRGD---- 339
QY 372 FSAVQVKNVYENIKGTSATKVAIKFDCSTNPFCEGILIMENITLV--VGSEG---KPEE-- 424
DB 340 -SQVAISDILFKNIRGTTIKDVQVIMKSKSVDCQGVNVADVNDLVYKTKGGERKSSGG 398
QY 425 ---ATCKNVH 431
DB 399 LVGALCDNAN 408

RESULT 7
PG1L PRUPE STANDARD; PRT; 393 AA.
ID PG1L PRUPE
AC P48979;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase).
OS Prunus persica (Peach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
```

```

OC eutrosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=3760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Flavourcrest;
RX MEDLINE=94302167; PubMed=8029352;
RA Lester D.R., Speaters J., Orr G., Brady C.J.;
RT "Peach (Prunus persica) endopolygalacturonase cDNA isolation and mRNA
  analysis in melting and nonmelting peach cultivars.";
RL Plant Physiol. 105:225-231(1994)
CC -1- FUNCTION: Acts in concert with the pectinesterase, in the ripening
  process. Is involved in cell wall metabolism, specifically in
  polyuronide degradation.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
  galactosiduronic linkages in pectate and other galacturonans.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76735; CAAS4150.1; -.
DR PIR; S40123; S40123.
DR HSSP; P26509; 1BHE.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR006626; Pbh1.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR SMART; SM00710; Pbh1; 6.
DR PROSITE; PS00502; POLYGALACTURONASE; 1.
KW Hydrolyase; Glycosidase; Cell wall; Signal; Fruit ripening;
KW Glycoprotein.
FT SIGNAL 1 25
FT CHAIN 26 393
FT ACT SITE 242 242
FT CARBOHYD 260 260
SQ SEQUENCE 393 AA; 41472 MW; 53DCC6944D22BF9 CRC64;
  N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match
Best local similarity 28.8%; Score 687.5; DB 1; Length 393;
Matches 148; Conservative 60; Mismatches 137; Indels 13; Gaps 8;

QY 77 NVLSFGAKGDKTYDNTIAFEQANMEACSSRTPVOFVVKNNKYLKQTFSGPCR-SSTS 135
  29 NVASLGAADKDTSTRAFLISAMAKACSMNPGVITYYPAG-TFFLRDVVFGSPCKNAIT 87
  136 VKIFGSLAASKISDYK---DRRLMIAPDSYQNLVVGGGGTINNGQVWMSCKINKSL 192
  88 FRINGTIVP---SDRYVIGNAANWIPFHHVNGVTI-SGGLDDQGTLMW--ACKACHGE 141
  193 PCRDAPALTFWNCNKLKVNLSKSKNAQOIHKESSCTNVVNASLMTNASKSPNTGVH 252
  142 SCPEGATTLGFSDSNNIYVSGIASLNSQMFHIVINDFQNVOMQGVRSRSGNSPNTDGIH 201
  253 VSNQVYQIISTIIIGTGDPCISIVSGSNVQATNITGCPHGHSISGLSGNSRAYSVNV 312
  202 VQWSSGYVTLMSKATGDCVSIIPGTSNMIIEGACGPGHISISLGEKEQEBRAGVQNV 261
  313 TVNEAKTIIGENVRIKTW--QGSGSGAASNIKELNVEMODVYKPIIIDQNYCDRVEPCIQ 371
  262 TVKVTSTSGTNGRIKISWGRPSTGFPANILFGHATMNWVNPFIIVIOHYPEDKKCGGQ 321
  372 FSAVQVKNVVENIKGTSATKVAIKPDCSTNFPCEGIIMENINLVSGSGKSEATCKN 429
  322 VSGVQISDVYEDIHGTSATEVAVFDCSPGHPREIKLEDKLTYNK-CAASSCSH 378
  RESULT 8
  PGLR_GOSHI

```

```

ID PGLR_GOSHI STANDARD; PRT; 407 AA.
AC Q39786;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polygalacturonase precursor (BC 3.2.1.15) (PG) (Pectinase).
GN G9.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eutrosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Coker 312; TISSUE=pollen;
RX MEDLINE=95161720; PubMed=7858233;
RA John M.E., Petersen M.W.;
RT "Cotton (Gossypium hirsutum L.) pollen-specific polygalacturonase
  mRNA: tissue and temporal specificity of its promoter in transgenic
  tobacco.";
RL Plant Mol. Biol. 26:1989-1993(1994).
CC -1- FUNCTION: May function in the depolymerization of the pectin in
  its walls during pollen tube elongation, or in that of the pistil
  during pollination.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
  galactosiduronic linkages in pectate and other galacturonans.
CC -1- TISSUE SPECIFICITY: Pollen.
CC -1- DEVELOPMENTAL STAGE: Appears 12 days before anthesis and maximum
  levels are seen in pollen on the day of anthesis.
CC -1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09717; AAA82167.1; -.
DR PIR; S52006; S52006.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR006626; Pbh1.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR SMART; SM00710; Pbh1; 5.
DR PROSITE; PS00502; POLYGALACTURONASE; 1.
KW Hydrolyase; Glycosidase; Cell wall; Signal; Glycoprotein.
KW Glycoprotein.
FT SIGNAL 1 26
FT CHAIN 27 407
FT ACT SITE 244 244
FT CARBOHYD 182 182
FT CARBOHYD 272 272
FT CARBOHYD 272 272
FT CARBOHYD 302 302
FT CARBOHYD 331 331
SQ SEQUENCE 407 AA; 43921 MW; B81B2BC4C312D195 CRC64;
  N-LINKED (GLCNAC. . .) (POTENTIAL).
  N-LINKED (GLCNAC. . .) (POTENTIAL).
  N-LINKED (GLCNAC. . .) (POTENTIAL).
  N-LINKED (GLCNAC. . .) (POTENTIAL).
  N-LINKED (GLCNAC. . .) (POTENTIAL).
  N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match
Best local similarity 27.9%; Score 667; DB 1; Length 407;
Matches 140; Conservative 69; Mismatches 150; Indels 6; Gaps 5;

QY 67 KVDKGIKVINVLVSFGAKGDKTYDNTIAFEQANMEACSSRTPVOFVVKNNKYLKQTF 126
  23 KVQSDARDV--VAFAFGAKADKDTLSKPFUDAWKACASVTPVIVLPKG-TYLLSKNLT 79
  127 SGPGRSSISVYKIFSLAASKISDYKDRRLMIAPDSYQNLVVGGGGTINNGQVWMSBSC 186
  80 EGPCKARIEINVOQTIIQAPADPSAFKDPN-WRFYSVENFPMFGGJFDGGSTAYEKNT 138
  187 KINKSLPCRDAPALTFWNCNKLKVNLSKSKNAQOIHKESSCTNVVNASLMTNASKSP 246
  139 CENRPRSK-LPVNIRPFLTNALIIDTISKDSKLFHINVFACKNITLERKIEAPDESP 197

```



Db	159	LNKNAIFEDGGGSLAMKANDCA--KTGKCNLSPIINRFGTNSKINSITSTNSKLFHNM	216
QY	226	FESCINVAASNLMTNINASAKSPRTDSDVHNSNTPCYQISDIIITGDPCISIVSSQNVQAT	285
Db	217	ILNCNITLSDIGDAPPELSLTDIIHIGRSSGVMIIQAKITKTDGDCVSIIGDTEMLIVE	276
QY	286	NITCGPGHGISIGSLGSGNSEAYVENVNEMAKIIGAEAGVRILKTWQGG-SCGASNIKFL	344
Db	277	NVEGGPGHGISIGSLGRYPNEGPRVAGVIVRKCLIKNTDNGVARIKTPGSGPRGLASIIIE	336
QY	345	NVEMQVKYPIIIIDQNYCDREVEPCIQGFSAYGVKNVYENIKGTSATKVAIKFDCSTNPP	404
Db	337	DITMDNVAFLPYLIDDEYCPYHGCKAGVPSQVKLSLDTVIIGIKISTSATKYAVLKMCKSGYV	396
QY	405	CGEIIEMENINLY--GESGKPSFATCKNV	430
Db	397	CTNIALSDIINLVHNGKEG-PAVASASNI	423

RESULT 10	PGLR_GOSBA	STANDARD;	PRT;	407 AA.
ID	PGLR_GOSBA			
AC	Q39766;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Polysialacturonase precursor (EC 3.2.1.15) (PG) (Pectinase) .			
GN	G9.			
OS	Gossypium barbadense (Sea-island cotton) (Egyptian cotton) .			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eucosids .II; Malvales; Malvaceae; Malvoideae; Gossypium.			
OX	NCBI_TaxID=3634;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Pollen;			
RX	MEDLINE=95161720; PubMed=7858233;			
RA	John M.E., Petersen M.W.;			
RT	"Cotton (Gossypium hirsutum L.) pollen-specific polysialacturonase			
RT	mRNA: tissue and temporal specificity of its promoter in transgenic			
RT	tobacco";			
RL	Plant Mol. Biol. 26:1989-1993(1994) .			
CC	-I- FUNCTION: May function in the depolymerization of the pectin in			
CC	its walls during pollen tube elongation, or in that of the pistil			
CC	during pollination.			
CC	-I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-			
CC	galactosiduronic linkages in pectate and other galacturonans.			
CC	-I- TISSUE SPECIFICITY: Pollen (By similarity) .			
CC	-I- SIMILARITY: Belongs to family 28 of glycoyl hydrolases.			
CC	-----			
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/isb-sib.ch">http://www.isb-sib.ch/announce/isb-sib.ch</a> ).			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U09805; AAA58322.1; -			
DR	InterPro; IPR000743; Glyco_hydro_28.			
DR	InterPro; IPR006626; PBH1.			
DR	Pfam; PF00295; Glyco_hydro_28; 1.			
DR	SMART; SM00710; PBH1; 5.			
DR	PROSITE; PS00502; POLYGALACTURONASE; 1.			
KW	Hydrolase; Glycosidase; Cell wall; signal; Glycoprotein.			
FT	SIGNAL	1	26	POTENTIAL.
FT	CHAIN	27	407	POLYGALACTURONASE.
FT	ACT SITE	244	244	BY SIMILARITY.
FT	CARBOHYD	182	182	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	267	267	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	272	272	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	302	302	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	331	331	N-LINKED (GLCNAC . . .) (POTENTIAL) .

SQ SEQUENCE 407 AA; 43943 MW; EBE842A77158CCF0 CRC64;

Query Match	27.9%	Score 666;	DB 1;	Length 407;
-------------	-------	------------	-------	-------------

Matches 140; Conservative 69; Mismatches 150; Indels 6; Gaps 5;

67 KVDKNGIKVINLVSFGAKGDKTYDNIAFEQAWNEACSSRTPVQFVVPKNKNYLLKQITF 126

Db 23 KVQPDADFV--VAKFGAKADGKTIDLSKPFIDAWKEACASVTPSTVVIPKG-TYLLSKVNL 79

QY 127 SGPGRSSISVKIFGSLAASSKISDYKDRRLWIAFDVQNLVGGGTINGNGQVWWPSSC 186

Db 80 EGPCKAPIEINVGTIQAPADPSAFKDPN-WVRFYSVENFKMFGGGIFDGGGSIAYEKNT 138

QY 187 KINKSLPCRDAPTALTFWNCKNLKVNNLKSNAQQIHKEFECTNVVASNLMINASAKSP 246

Db 139 CENREFRSK-LPVNIRFDEVTNALI QDITSKSKL FHINVFACKNITLERLKIAPDESP 197

247 NTDGVHVSNTQYIQISDTIIIGTGDDCISIVSGSQNVQATNITCGPGHGISIGSLGSGNSE 306

Db 198 NTDGIHMGKSEGVNIASDIKTGDDCISIGDGTKNMVIKEITCGPGHGISIGSLGKFQNE 257

307 AYVSNVTNEAKIIGAENGVRIKTWQGGSGQA-SNIKFLNVEMQDVKYPPIIDQNYCDRV 365

Db 258 EPVEGIKISNCTITNTSNGARIKTMPEGEHGAVSEIHFEDITMNNVSSPILIDQYCPWN 317

QY 366 EPCIQFSAVQKNNVYENIKGTSATKVAIKFDCSTNFPCEGIIMENINLVGESGKPSEA 425

Db. 318 KCKKNEESKVL SNISFKNIRGTSALPEAIKFI CS GSSPCQVELADIDIDIQNGAEPATS 377

QY 426 TCKNV 430

Dp 378 QCLNV 382

RESULT 11

PGLT MAIZE	
ID PGLT MAIZE	STANDARD.
	BPT. 410 AA

DT 01-JUN-1994 (Re) 29 (Created)

DT	01-JUN-1994 (Rel. 29, Last sequence update)
BT	10-OCT-2003 (Rel. 42, Last annotation update)

DE Exopolysaccharonase precursor (EC 3.2.1.67) (ExoPG) (Pectinase)

GN PG2C.  
OS Zea mays (Maize)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta: Magnoliopsida: Poales: Poaceae:

OC PACCBAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI TaxID=4577.

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=93164262; PubMed=8433375;  
RC STRAIN=cv; Missouri 17; TISSUE=Leat;  
S

KA Barakate A., Martin W., Quigley F., Macne R.;  
 BT "Characterization of a multigene family encoding an

RT exopolysaccharonase in maize.";  
J. Mol. Biol. 229:797-801 (1993).

CC -1- FUNCTION: May function in depolymerizing pectin during pollen development, germination, and tube growth. Acts as an exo-

CC -1- CATALYTIC ACTIVITY: { (1,4)-alpha-D-galacturonide } (N) + H(2)O =  
CC polygalacturonase;

```
cc      { (1,4) -alpha-D-galacturonide } (N-1) + D-galacturonate.
cc      -|- SUBCELLULAR LOCATION: Secreted.
```

cc -1- TISSUE SPECIFICITY: Pollen.  
cc -1- DEVELOPMENTAL STAGE: Late stages of pollen development.

```

cc      -|- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES.
cc      -----

```

between the Swiss Institute of Bioinformatics and the EMBL outstation - CC

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modelled and this statement is not removed. usage by and for commercial





RX	MEDLINE=92032781; PubMed=1932692;
RA	Nigorete M.F., Dubald M., Mandaron P., Maché R.;
RC	"Characterization of pollen polygalacturonase encoded by several cDNA clones in maize.";
RT	Plant Mol. Biol. 17:1155-1164(1991).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN=cv. Missouri 17, TISSUE=leaf;
RC	MEDLINE=93164262; PubMed=8433375;
RA	Barakate A., Martin W., Quigley F., Maché R.;
RX	MEDLINE=93004490; PubMed=1391780;
RT	Allén R.L., Lonsdale D.M.;
RL	"Sequence analysis of three members of the maize polygalacturonase gene family expressed during pollen development.";
PL	Plant Mol. Biol. 20:343-345(1992).
RN	[4]
RP	SEQUENCE OF 1-306 FROM N.A.
RX	MEDLINE=94035141; PubMed=8106080;
RA	Allén R.L., Lonsdale D.M.;
RT	"Molecular characterization of one of the maize polygalacturonase gene family members which are expressed during late pollen development.";
PL	Plant J. 3:261-271(1993).
CC	-I- FUNCTION: May function in depolymerizing pectin during pollen development, germination, and tube growth. Acts as an exo-polygalacturonase.
CC	-I- CATALYTIC ACTIVITY: {(1,4)-alpha-D-galacturonide}(N) + H(2)O = {(1,4)-alpha-D-galacturonic acid}(N-1) + D-galacturonate.
CC	-I- SUBCELLULAR LOCATION: Secreted.
CC	-I- TISSUE SPECIFICITY: Pollen.
CC	-I- DEVELOPMENTAL STAGE: Late stages of pollen development.
CC	-I- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X57627; CAA40850.1; .-
DR	EMBL; X57628; CAA40851.1; ALT INIT.
DR	EMBL; X57743; CAA40910.1; .-
DR	EMBL; X57743; CAA40910.1; .-
DR	EMBL; X65844; CAA46679.1; .-
DR	EMBL; X65845; CAA46680.1; .-
DR	EMBL; X62384; CAA44248.1; .-
DR	EMBL; X62385; CAA44249.1; .-
DR	EMBL; X66692; CAA47234.1; .-
DR	PIR; S18570; S18570.
DR	PIR; S25824; S25824.
DR	PIR; S30064; S30064.
DR	MaizeDB; 25864; .-
DR	InterPro; IPR000743; Glyco_hydro_28.
DR	InterPro; IPR006626; PDH1.
DR	Pfam; PF00295; Glyco_hydro_28; 1.
DR	SMART; SM00710; PbH1; 5.
DR	PROSITE; PS00502; POLYGALACTURONASE; 1.
KW	Hydrolase; Glycosidase; Cell wall; Signal; Glycoprotein;
KW	Multigene family.
FT	SIGNAL 1..22
FT	CHAIN 23..410
FT	ACT SITE 256..256 EXOPOLYGALECTURONASE.
FT	CARBHYD 89..89 PROBABLE.
FT	CARBHYD 246..246 N-LINKED (GLCNAC...)(POTENTIAL).
FT	CARBHYD 349..349 N-LINKED (GLCNAC...)(POTENTIAL).
FT	CARBHYD 349..349 N-LINKED (GLCNAC...)(POTENTIAL).

FT	CARBONYD	387	387	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	359	359	A -> V (IN PGSL14).
SEQ	SEQUENCE	410 AA;	43443 MW;	0A6E779644D818BA CRC64;
Query Match				
Best Local Similarity 34.4%; Pred. No. 7.6e-37; Length 410;				
Matches 134; Conservative 77; Mismatches 157; Indels 22; Gaps 8;				
Qy	66	DKVDKNGIKV-----INVLSPGAKGDKTYDNIAPFQAMNEACSSRRPVPFVPEPKK	117	
Db	23	EKESSKGIIDAKASGPGGSPDITKLGASGNGKTSTKAVGAMASACGGTGKQTILIPKG-	81	
Qy	118	NYLLKQTFPGSPGSSISVFKIFGLSEASSKSIDYKDRRLIAPFSVQNLVYGGGTTNGN	177	
Db	82	DFIVGQNLFTFGPGCKGVDVITQVDDGLATPTDLSQYKOHGMNIETILRDVNLVITGKNLDGQ	141	
Qy	178	G-QYVWVSSCKINSLPCRDAPPTLTFWNCNKLKVNLLKSKNAQOHHKFECSNNVASN	236	
Db	142	GPAVMSKNSC--TKYTCCKILPNSLVMDPVNNGEVSQVTLNLSKFFPMNMYRCKDMLIKD	199	
Qy	237	LMINASAKSPNTDGVHVSNTQYIQISDPTIIGTDDCISIVSGSONVQATNITCGPGHGIS	296	
Db	200	VTVLAPDPSPTDQIHMGDSGGITTTNTVIGVGDDCISIGPSTSKVNTIGVTCGPGHGIS	259	
Qy	297	IGSISGSGNSFAYSNVTYVNAEKIIGANGVIRIKTWQCGSG--QASNIKFLVEMQDYKYP	354	
Db	260	IGSLGRYKDEBDVTDINVKDCTLTKTFEVRIRIAXEYDPAASVLTYSKLYHENIKMEDSAMP	319	
Qy	355	IIDQNTCDREVEPICQFSAVQVKNVYENIKGTSATKAVIKPDCSTNPCEGIMENIN	414	
Db	320	IFIDMKCPNKLCTANASKVTYKDYTFKNTITGTSIFPEAVSLCTAKVPCGTVMDDVN	379	
Qy	415	L--VGESGKPEATCKNVHFNNAEHVTPHC	442	
Db	380	VEYSGTNNK-TMALC-----TNKKGSTKGC	403	
RESULT 14				
PGLS MAIZE				
ID	PGLS MAIZE	STANDARD;	PRT;	410 AA.
AC	P3538;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Exopolysaccharide precursor (EC 3.2.1.67) (ExoPG) (Pectinase)			
DE	(Galacturan 1,4-alpha-galacturonidase).			
GN	PG9.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	PACCAD clade; Panicoideae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Missouri 17; TISSUE=leaf;			
RX	MEDLINE=33164262; PubMed=8433375;			
RA	Barakate A., Martin W., Quigley F., Macho R.;			
RT	"Characterization of a multigene family encoding an			
RT	exopolysaccharurone in maize.";			
RU	J. Mol. Biol. 229:797-801(1993).			
CC	-1- FUNCTION: May function in depolymerizing pectin during pollen			
CC	development, germination, and tube growth. Acts as an exo-			
CC	polysaccharuronase.			
CC	-1- CATALYTIC ACTIVITY: {(1,4)-alpha-D-galacturonide}(N) + H(2)O =			
CC	{(1,4)-alpha-D-galacturonide}(N-1) + D-galacturonate.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- TISSUE SPECIFICITY: Pollen.			
CC	-1- DEVELOPMENTAL STAGE: Late stages of pollen development.			
CC	-1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X64408; CAA45751.1; -  
DR PIR; S30066; S30066.  
DR MaizeDB; 25864; -  
DR InterPro; IPR000743; Glyco\_hydro\_28.  
DR InterPro; IPR006626; Pbh1.  
DR Pfam; PR00295; Glyco\_hydro\_28; 1.  
DR SMART; SM00710; Pbh1; 5.  
DR PROSITE; PS00502; POLYGLACTURONASE; 1.  
KW Hydrolyase; Glycosidase; Cell wall; Signal; Glycoprotein;  
KW Multigene family.  
FT SIGNAL 1 22  
FT CHAIN 23 410 EXOPOLYGLACTURONASE.  
FT ACT SITE 236 256 PROBABLE.  
FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 410 AA; 43415 MW; 942153CB50DEABA CRC64;

Query Match 26.2%; Score 626; DB 1; Length 410;  
Best Local Similarity 34.6%; Pred. No. 8.9e-37;  
Matches 135; Conservative 75; Mismatches 158; Indels 22; Gaps 8;

QY 66 DAVDKNKGIKV-----INVLSFGAKGDKGTVDNIAFEQAMNEACSSRTPVQFVPPKXK 117  
DB 23 EKESKGIAPAKASGPGSGFDITRLGASGNGKTDSTVAQVAMASACGGTGKOTILLPKG- 81  
QY 118 NVLLKQITPSGPRSSIVKIFGSLFASSKISDYKDRRLMTAPDSVQNLVVGSGGTINCN 177  
DB 82 DFLVGLNNTGPKGKGVTTIQVGNLTLATDLSQYKHEGMIETLRDNLVITKGMDDQ 141  
QY 178 G-QVWMPSSCKINKSLPCRDAPLALTFMNCNKLKVNLSKNAQOIHIFESECTNVASN 236  
DB 142 GAVNMGKNSC--TKKYDCKILPNSLYMDFVNNGEVSGVTLNKKFHHMNVYQCKNMLIKD 199  
QY 237 LMINASAKSPNTDGVAVNSTQYIQISDTIIGTDDCISVSGSQNVQATNITGPGHGHS 296  
DB 200 VVTLAPGDSPTNDGIMKDSGSLITNTYIVGAGDDCISIGPGSKVNIIGVGTGPGHGHS 259  
QY 297 IGSLSGNSGSEAVSVNTVNEAKIIGAENGVRITKMOGSGS--QASNIKFLNVEQDYKYP 354  
DB 260 IGSLSGNYKBEKDYTDINVKDCITTKTMEFVRIRAYEDASVLTFSKIHENIMEDSAND 319  
QY 355 IITDQNYCDRVBPCIOQFSNAVQKNVYENIKGTSATKVAIKFDGSTNPPCEGIIMENTN 414  
DB 320 IFTDMKCPNKLCTANGASKVTVKDYTFKNTIGTSSTPKALISLCTAKVPCGTATMDVNV 379  
QY 415 L--VGSSEKPESEATCKNVAFNNAEHVTPHC 442  
DB 380 VEYSGTNMK-TVAIC-----TNAGSGTKGC 403

RESULT 15  
PGLR\_MEDSA  
ID PGLR\_MEDSA STANDARD; PRT; 421 AA.

AC Q40312;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Polyalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase).  
OS Medicago sativa (Alfalfa).  
OC Burkholderia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid1; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.  
OX NCBI\_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. C2-4; TISSUE=Pollen;

RA Qiu X., Erickson L.; to the EMBL/GenBank/DBJ databases.  
RL Submitted (MAR-1995)  
CC -1- FUNCTION: May function in the depolymerization of the pectin in  
CC its walls during pollen tube elongation, or in that of the pistil  
CC during pollination (By similarity).  
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-  
CC galactosiduronic linkages in pectate and other galacturonans.  
CC -1- TISSUE SPECIFICITY: Pollen specific.  
CC -1- SIMILARITY: Belongs to family 28 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U20431; AAA62286.1; -  
DR PIR; T09398; T09398.  
DR InterPro; IPR000743; Glyco\_hydro\_28.  
DR InterPro; IPR006626; Pbh1.  
DR Pfam; PR00295; Glyco\_hydro\_28; 1.  
DR SMART; SM00710; Pbh1; 4.  
DR PROSITE; PS00502; POLYGLACTURONASE; 1.  
KW Hydrolyase; Glycosidase; Cell wall; Signal; Glycoprotein.  
FT SIGNAL 1 20  
FT CHAIN 21 421 POLYGLACTURONASE.  
FT ACT SITE 242 242 BY SIMILARITY.  
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 265 265 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 421 AA; 43953 MW; 6E1599E2BD9C73DE CRC64;

Query Match 24.7%; Score 589.5; DB 1; Length 421;  
Best Local Similarity 33.6%; Pred. No. 3.4e-34;  
Matches 128; Conservative 68; Mismatches 136; Indels 49; Gaps 11;

QY 75 VINVLISFGAKGDKGTVDNI--AFEQAMNEACSSRTPVQFVPPKKNYLLKQITPSGPRS 132  
DB 24 VLDISKFG-----CKRPSDGLQALTSAMNEACSTTAKIVIPAG-TYQNGIIEIKGCKA 78  
QY 133 SISVKIIEGLAESKISDYKDRRLMTAPDSVQNLVVGSGGTINGNQVWMPSSCKINKSL 192  
DB 79 PIELQVDGTIOAPADPSVAKTEQWEPFLYMDHLTLISGKGVFPGQATVY-----KXAA 132  
QY 193 PCRDAPLALTFMNCNKLK-----VNN-----LKSNAQOIHIFESECTNVASN 236  
DB 133 PA-----SANGKNSKNSKVENNFGFNFNNSIVRGVTSKOSKFNHVMVFGCKNITFDG 185  
QY 237 LMINASAKSPNTDGVAVNSTQYIQISDTIIGTDDCISVSGSQNVQATNITGPGHGHS 296  
DB 186 FTITAPGDSPTNDGIMKDSGVKILNTNIGTDDCVSLGDSKQITVCGSVNGPGHGHS 245  
QY 297 IGSLSGNSGSEAVSVNTVNEAKIIGAENGVRITKMOGSGS--QASNIKFLNVEQDYKYP 354  
DB 246 VGSLSGFTTEENVEGITVKKCTLTANDNGVRIRKTDAPOTITVSDIHFEIDIMTNVKNP 305  
QY 355 IITDQNY-----CDREPCIOQFSNAVQKNVYENIKGTSATKVAIKFDGSTNPPCEGI 409  
DB 306 VITDQYYPWNOQSKNP-----SKTKLSKFSKKNVIGSIAEGVVLICSSAVPCDQVE 360  
QY 410 MENINIVGSGSEKPESEATCKNV 430  
DB 361 LANNVDL-KENGAETTAKCTNV 380

Search completed: July 21, 2004, 17:06:51  
Job time : 23.6574 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: July 21, 2004, 17:02:32 ; Search time 69.4797 Seconds  
(without alignments)  
2075.310 Million cell updates/sec

Title: US-10-018-604-2

Perfect score: 2390  
Sequence: 1 MWIQNSILLIIPASSIS.....VTPHCTSLSEDEALVNY 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SEPREMBL\_25:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteriap:\*  
17: sp.\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	53.6	463	10	Q9M6S2 actinidia c
2	1204.5	50.4	444	10	Q94B15 vitis vinif
3	1185.5	49.6	428	10	Q944B5 glycine max
4	1185	49.6	438	10	Q95FB7 arabidopsis
5	1149	48.1	433	10	Q8RY29 arabidopsis
6	1130	47.3	434	10	Q8YDY7 brassica na
7	1128	47.2	461	10	Q81246 cucumis mel
8	1127	47.2	431	10	Q23147 arabidopsis
9	1115.5	46.7	426	10	Q22935 arabidopsis
10	1113	46.6	433	10	Q42399 brassica na
11	1108	46.4	458	10	Q43063 prunus pers
12	1106	46.3	433	10	Q8VXT3 brassica ra
13	1103	46.2	433	10	Q42636 brassica na
14	1073	44.9	460	10	Q8GTP8 pyrus commu
15	1061.5	44.4	335	10	Q85886 rhus idaeu
16	1034.5	43.3	405	10	Q84ZP3 oryza sativ

17	1022.5	42.8	422	10	Q84ZP5 oryza sativ
18	1014	42.4	444	10	Q84ZN9 oryza sativ
19	988.5	41.4	429	10	Q84M40 oryza sativ
20	925	38.7	452	10	Q9M7D3 lycopersico
21	914	38.2	444	10	Q94AJ5 arabidopsis
22	898.5	37.6	459	10	Q9SSC2 arabidopsis
23	840.5	35.2	468	10	Q9CAL5 arabidopsis
24	825.5	34.5	435	10	Q9SLP3 cucumis sat
25	789	33.0	514	10	Q8H987 cryptomeria
26	788.5	33.0	514	10	Q8H989 cryptomeria
27	788	33.0	457	10	Q9ZUR7 arabidopsis
28	787.5	32.9	434	10	Q9FXC1 arabidopsis
29	787.5	32.9	514	10	Q8H988 cryptomeria
30	764	32.0	507	10	Q9FY19 juniperus a
31	749.5	31.4	1161	10	Q9IOD1 arabidopsis
32	741.5	31.0	491	10	Q9FWX5 arabidopsis
33	736	30.8	423	10	Q81245 cucumis mel
34	734	30.7	503	10	Q94IT1 oryza sativ
35	730	30.5	392	10	Q22311 lycopersico
36	728	30.5	443	10	Q9SWS3 glycine max
37	728	30.5	405	10	Q22818 arabidopsis
38	727	30.4	392	10	Q40135 lycopersico
39	726	30.4	392	10	Q96487 lycopersico
40	719.5	30.1	398	10	Q8GTP7 pyrus commu
41	718.5	30.1	395	10	Q22610 lycopersico
42	718	30.0	351	10	Q84N55 prunus pers
43	718	30.0	435	10	Q91YJ5 arabidopsis
44	715.5	29.9	397	10	Q9ZT65 lycopersico
45	715	29.9	392	10	Q80559 arabidopsis

## ALIGNMENTS

RESULT 1  
ID Q9M6S2 PRELIMINARY; PRT; 463 AA.  
AC Q9M6S2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Polygalacturonase A.  
GN PGA.  
OS Actinidia chinensis (Kiwi) (Yangtiao).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Ericales; Actinidiaceae; Actinidia.  
OX NCBI\_Taxid=3625;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=pericarp;  
RX MEDLINE=20252524; PubMed=10794531;  
RA Wang Z.Y., MacRae F.A., Wright M.A., Bolitho K.M., Ross G.S.,  
Atkinson R.G.;  
RT "Polygalacturonase gene expression in kiwifruit: relationship to fruit  
softening and ethylene production.";  
RL Plant Mol. Biol. 42:317-328(2000).  
CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
(POLYGALACTURONASES).  
DR EMBL; AF152758; AAF71160.1; -;  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0004650; F:polygalacturonase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR000743; Glyco\_hydro\_28.  
DR InterPro; IPR006626; PDB1.  
DR InterPro; IPR000408; Reg\_chr\_condens.  
DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
DR SMART; SM00710; PDB1; 4.  
DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
DR PROSITE; PS00626; RCCL\_2; 1.  
KW Cell wall; Glycosidase; Hydrolase.  
SQ SEQUENCE 463 AA; 50352 MW; 7EBD7EEFE54988FA CRC64;

Query Match 53.6%; Score 1282; DB 10; Length 463;  
 Best Local Similarity 54.2%; Pred. No. 6.9e-76;  
 Matches 256; Conservative 72; Mismatches 114; Indels 30; Gaps 8;

QY 1 MTIGRN--SILLIIIFASSISICRSNVIDNLFKQVYNDLIEOFBAPQALSTYLSKN 58  
 Db 1 MALQRHFQFQVITLLPFIITLPGTSAVHED---PRHDYHLE-BYGVDFRAYSYITTI 55  
 QY 59 IESNNNDKVDKNGI-----KVINVLSPKAGDGKTYNDIAEQWANE 101  
 Db 56 GDNFGSSMSHENEIPGLRKVDYGDRLDASKTYNVDDFGAKDGDSD-DYKAFKXMKWA 114  
 QY 102 ACSSRTPVQVFPKKNKYLTKQITFGSGPCRSSISVKIFGSLFSSKISDY-KDRRLWIAF 160  
 Db 115 VCSSTSSANVLVPG-KYLVLPITFGPCCKSDLTMOIYGTLEASDDSDSKDGRHMLVF 173  
 QY 161 DSVQNLVVGGGGTINGNGQVWPPSSCKINKSLPCRDAPALTLPWNCNKLKYNLKSRYAQ 220  
 Db 174 DSVQNLVVGGGGTINGNGQVWPPSSCKINKSLPCRDAPALTLPWNCNKLKYNLKSRYAQ 220  
 QY 221 QIHFKFSCCTNVASNLMINASAKSPNTDGVHVSNTQYIQISDTITIGTDDCISIVSGSQ 280  
 Db 234 QIHVSFPCNVAVQASNMVTTTPENSPNTDGIHVGTQINIHSSCVIEFGDDCISIVSGSQ 293  
 QY 281 NVQATNTTCGPHGISIGLSGNSSEAVSVNVTVNEAKTIIGANGVRIKTMQSGSGQASN 340  
 Db 294 KVRVNDITTCGPHGISIGLSGNSSEAVSVNVTVNEAKTIIGANGVRIKTMQSGSGQASN 353  
 QY 341 IKFLVEMQDVKYPIIIDONYCDREVPICIQFSAVOYKVVYENIKGTSATKVAIKFDCS 400  
 Db 354 IKFQNVEMHNVENPIIIDONYCDREVPICIQFSAVOYKVVYENIKGTSATKVAIKFDCS 413  
 QY 401 TNPFCGIIEMNTLVGSGKPSKSEATCKNVHFNNAEHVTPHCTSLSEDEA 452  
 Db 414 KRPFCQGIIVEDVLEIGGAARALCNVLESETGVVSPHCPD---GEEEDA 462

RESULT 2  
 Q94B15 PRELIMINARY; PRT; 444 AA.

AC Q94B15; TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Polylacturonase PGL  
 OS Vitis vinifera (Grape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC Vitaceae; Vitis.  
 CX NCBI\_Taxid=29760;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Shiraz.  
 RA Numan K.J., Davies C., Robinson S.P., Fincher G.B.;  
 RT "Characterization of cell wall modifying enzyme activities and their  
 RT corresponding cDNAs during grape berry development.",  
 RL Planta 0:0-0(2001)  
 CC -|- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 CC EMBL; AY043233; AAK81876.1; -.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0004650; F:polylacturonase activity; IEA.  
 DR GO; GO:0005975; F:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000743; Glyco\_hydro\_28.  
 DR InterPro; IPR006626; Pbh1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR SMART; SM00710; Pbh1, 4.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 DR PROSITE; PS00626; RCG1\_2; 1.  
 KW Cell wall; Glycosidase; Hydrolase.

SQ SEQUENCE 444 AA; 47419 MW; D94259F4792F132 CMC64;  
 Query Match 50.4%; Score 1204.5; DB 10; Length 444;  
 Best Local Similarity 61.5%; Pred. No. 7.7e-71;  
 Matches 228; Conservative 53; Mismatches 87; Indels 3; Gaps 3;

QY 73 IKVINVLSPKAGDGKTYNDIAEQWANEACSSRTPVQVFPKKNKYLTKQITFGSPCRS 132  
 Db 75 VKMVNVNITGAKKGGSDATF-AFKAKAKACSSRQSV-LVFPKKNKYLTKQITFGSPCRS 132  
 QY 133 SISVKIFGSLFSSKISDY-KDRRLWIAFDSQNLVVGGGGTINGNGQVWPPSSCKINKS 191  
 Db 133 SITVQIVGVQASIDRSAYSNDMTHMLIFENVQNLAVQGGGTINGNGKTMWNSCKVNYD 192  
 QY 192 LPCRDAPALTLPWNCNKLKYNLKSRYAQQIHFKFSCCTNVASNLMINASAKSPNTDGV 251  
 Db 193 LPCRDAPALTLPWNCNKLKYNLKSRYAQQIHFKFSCCTNVASNLMINASAKSPNTDGI 252  
 QY 252 HVSNTQYIQISDTITIGTDDCISIVSGSONVQATNITCGPHGISIGLSGNSSEAVSVN 311  
 Db 253 HVSNTQYIQISDTITIGTDDCISIVSGSONVQATNITCGPHGISIGLSGNSSEAVSVN 312  
 QY 312 VTVNEAKTIIGANGVRIKTMQSGSGQASNIKFLVEMQDVKYPIIIDONYCDREVPICIQ 371  
 Db 313 VTVNGATLSTGTNGVRIKTMQSGSGQASNIFKQNVEMHNVENPIIIDQYCDQSKPCSQ 372  
 QY 372 FSAVOYKVVYENIKGTSATKVAIKFDCSTNPFCGIIEMNTLVGSGKPSKSEATCKNVH 431  
 Db 373 SSAVOYKVVYENIKGTSATKVAIKFDCSTNPFCGIIEMNTLVGSGKPSKSEATCKNVH 432  
 QY 432 FNNAEHTVPHC 442  
 Db 433 VTVMGDVSPNC 443

RESULT 3  
 Q94B5 PRELIMINARY; PRT; 428 AA.

AC Q94B5; TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Denilence-related endopolylgalacturonase.  
 GN SDPG.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC Eufosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 CX NCBI\_Taxid=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Williams 82.  
 RA Christiansen L.C., Dal Degam F., Ulyskov P., Borkhardt B.;  
 RT "Examination of the denilence zone in soybean pods and isolation of a  
 RT denilence-related endopolylgalacturonase gene.",  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -|- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 CC EMBL; AF34714; ALU30418.1; -.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0004650; F:polylacturonase activity; IEA.  
 DR GO; GO:0005975; F:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000743; Glyco\_hydro\_28.  
 DR InterPro; IPR006626; Pbh1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR SMART; SM00710; Pbh1, 6.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 428 AA; 45830 MW; F553DBAC29762CD CRC64;

Query Match 49.6%; Score 1185.5; DB 10; Length 428;  
 Best Local Similarity 51.5%; Pred. No. 1.3e-69;

[illegible]







```

Db      168  INGNKTMWONCKIDSKPCTKAPALTLYNLKNLNVKNLRVKNAOQI0ISEKCNKE 227
QY      234  ASNMINVASAKPTDGVHVSNTQYIOISPTIIGTDDCISIVSGSNVQATNITCGPH 293
Db      228  VSNVEITAHGDSPTDGHITNTQNRVNSDSIDGTDCISIEDGTOLQFDLTCGPH 287
QY      294  GISIGLSGNSGSEAVSVNTVNEAKIIGAENGVRKTMWGSGGQASNIKELNEMODVKY 353
Db      288  GISIGLSGDNRSKAYVSGINVDGAKFSESDNGVRKITYGSSGTAKNIFKNINEMVKN 347
QY      354  PIIIDNVYCDRVEFCIOQFSNVQVKNVYENIKTSATKVAIKFDCSTNPFCEGIMENT 413
Db      348  PIIIDQYCK-DKCEDQESAVQVKNVYKNISGTSATDVAITLNCSEKYPQGIIVLENV 406
QY      414  NLVGESEKPEATCKNVHPNNAEHVPHCT 443
Db      407  KING-----GTASCKNANVKNQGVSPKCS 431

RESULT 9
ID      022935  PRELIMINARY;      PRT;      426 AA.
AC      022935;
DT      01-JUN-1998 (TREMBlrel. 05, Created)
DT      01-JUN-1998 (TREMBlrel. 05, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Putative polygalacturonase.
GN      AT2G41850.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_Taxid=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia;
RX      MEDLINE=20083487; Pubmed=10617197;
RA      Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA      Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA      Beell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Xoo H., Moffat K.S.,
RA      Cronin L.A., Shen M., Vanhaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA      Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA      Copenhaver G.P., Frazer D., Nierman W.C., White O., Eisen J.A.,
RA      Salzberg S.L., Fraser C.M., Venter J.C.;
RT      "sequence and analysis of chromosome 2 of the plant Arabidopsis
RT      thaliana.";
RL      Nature 402:761-768(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Columbia;
RA      Lin X.;
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC      -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC      (POLYGALACTURONASES).
EMBL; AC002339; AAC02763.1; -.
DR      PIR; H84846; H84846.
DR      GO; GO:0005618; C:cell wall; IEA.
DR      GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR      GO; GO:0004650; F:polygalacturonase activity; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR007043; Glyco_hydro_28.
DR      InterPro; IPR006626; PbH1.
DR      InterPro; IPR004081; Reg_chr_condens.
DR      Pfam; PF00295; Glyco_hydro_28; 1.
DR      SMART; SM00710; PbH1_5.
DR      PROSITE; PS00626; RCG1_2; 1.
DR      Cell wall; Glycosidase; Hydrolase.
SQ      SEQUENCE 426 AA; 45924 MW; F756754C0390A80B CRC64;

Query Match      46.7%; Score 1115.5; DB 10; Length 426;
Best Local Similarity 49.8%; Pred. No. 4; de-65;
Matches 222; Conservative 74; Mismatches 117; Indels 33; Gaps 7;

```

```

QY      6  NSIIILIIIFASSISTCRSNVIDNLFFKOYV-----DNILEQEFAPHPQVATLSKNIE 60
Db      6  NLVVFELMALIMSEWCKASRISPNVYDHSYKPKSLIRR--EDITGRSFRASLR 63
QY      61  SNNNIDRVKNGKIVINVLSEFAGKDGKTYDNIAFEQAMNEACSSRTVPQFVPRKNRYL 120
Db      64  TPT-----TVASVSPFGAKDGKDTDTQAFVNAWKKACSSNGAVNLVPRKNTYL 112
QY      121  LKQTFSGPGRSSISVYIFGSLSESSKISDYKDERLMTAPDSQMLVVGGS--GTINGG 178
Db      113  LKSTQLTGPQNSILTVQIFGLSLASQKRSYKDISKIMPDGVNLSVDDGDTGVVGNG 172
QY      179  QVWPPSSCKINKSLPCHDAPALTFFMNCNKLKVNLSKNAQOJHIFESCTNVASLIM 238
Db      173  ETWVQNSCKRKA-----KALTFYNSKSLIVNKLVRNNAQOJHISEKSNVQSVNV 225
QY      239  IMASAKSPNTDGVHVSNTQYIOISDTIIGTDDCISIVSGSNVQATNITCGPHGISIG 298
Db      226  VTAPADSPNTDGHITNTQNRVNSDSIDGTDCISIESGQNVQINDITCGPHGISIG 285
QY      299  SLGSGNSEAVSVNTVNEAKIIGAENGVRKTMWGSGGQASNIKELNEMODVKYPIID 358
Db      286  SLGDSNKAFFVSGVTVGAKLSTGDNVGRKITYGSGGTASNLIFQNDVKNPPIID 345
QY      359  QNYCDRVEPCIOQFSNVQVKNVYENIKTSATKVAIKFDCSTNPFCEGIMENTNLVGE 418
Db      346  QDYCDK-SKCTEKSANVQVKNVYRDISGTSASNAITFNKSNKYPQGIIVLRVNIKG- 403
QY      419  SKRSEATCKNVHPNNAEHVPHCTS 444
Db      404  -----GKATCTNANVVDKGAVLPCQNS 425

RESULT 10
ID      042399  PRELIMINARY;      PRT;      433 AA.
AC      042399;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Endo-POLYGALACTURONIDASE.
GN      PGAL.
OS      Brassica napus (Rape).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Brassica.
OX      NCBI_Taxid=3708;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. TOPAS;
RA      Sander L., Bottermann J., Ulvskov P., Borkhardt B.;
RL      Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC      -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC      (POLYGALACTURONASES).
EMBL; X98373; CA67020.1; -.
DR      EMBL; X98800; CA65072.1; -.
DR      GO; GO:0005618; C:cell wall; IEA.
DR      GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR      GO; GO:0004650; F:polygalacturonase activity; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR007043; Glyco_hydro_28.
DR      InterPro; IPR006626; PbH1.
DR      InterPro; IPR004081; Reg_chr_condens.
DR      Pfam; PF00295; Glyco_hydro_28; 1.
DR      SMART; SM00710; PbH1_6.
DR      PROSITE; PS00502; POLYGALACTURONASE; 1.
DR      PROSITE; PS00626; RCG1_2; 1.

```

KW Cell wall, Glycosidase, Hydrolase, Signal.  
SQ SEQUENCE 433 AA, 46595 MW, 1EDB29AC3B958260 CRC64;  
Query Match 46.6%; Score 113; DB 10; Length 433;  
Best Local Similarity 49.9%; Pred. No. 7.2e-65;  
Matches 223; Conservative 69; Mismatches 121; Indels 34; Gaps 8;  
QY 7 SILLIIIFASSISTCRSNVIDDNLFKQYVDNLEQEPFAHDFQAYLSTYLSKNIESNNMID 66  
DB 9 AVEFLCVLMLACCCQALLSNV-DDG-----YCHEDG---SFEESDLILKNDND 51  
QY 67 KY-----DKNGIKVINLSFGAKGDKTYDNTAFEOANWACSSRTFVQFVYPKKN 118  
DB 52 DVLTKSSDRPTTESSTVSFNGAKGDKTDTDQAFKKAMKACSTNGVTFLLPKGKT 111  
QY 119 YLLKQITSGPCRSISVYIFGSLFASSTISDY-KDRMIAPDSYONLVGG--GTIN 175  
DB 112 YLLKSIIRRGPKSIRSFQITGLTSLASTKSDSYNDKHMILLLEDVNNLSITDGSAGIVD 171  
QY 176 GNGQVWMPSSCKINKSLPCRDAPTLTFWNCNKLKYNLKSNAQOIHIKFEESTNVAS 235  
DB 172 GNGNIMWQNSCKIDSKPCTKAPTLTLNKLNVKNLRVANAQOIQISIEKNNVGVK 231  
QY 236 NLMINASAKSENTDGVHNSNTQYIQISDTTIGTDDCISIVSGSONVQATNITCGPHGI 295  
DB 232 NKLITAPGDSPTDGIHIVATKIRISNSDITGDDCISIEDGSONVQINDLTGPHGI 291  
QY 296 STSGSGNSSEAYVSNVTVNEAKIIGAENGVRKTMWGSGQASNTKFLNVEMODVKPI 355  
DB 292 STSGISGDDNSKAYVSGIDVGTLETEDGVALKTYGSGSGTRAKIKFNNIMDVKNPI 351  
QY 356 IIDQNYCDRVEPCIOQFSAVQKNNVYENIKGTSATKVAIKFDCSTNPFCEGIIMNTL 415  
DB 352 IIDQNYCDK-DKCEQGBASAVQNNVYQNIKGTSAITDAIMNCGVKPKQGIIVLENVI 410  
QY 416 VGESEKPESEATCKNVHFNNAHEVTHC 442  
DB 411 KG-----GKASCENNVKDKGTSPKC 432  
RESULT 11  
ID Q43063 PRELIMINARY; PRT; 458 AA.  
AC Q43063;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Polylacturonase (EC 3.2.1.15).  
GN PG.  
OS Prunus persica (Peach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
OC NCBI\_Taxid=3760;  
QY [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Maravilla;  
RA Lee E., Speirs J., Gray J., Brady C.J.;  
RT "Homologues to the tomato endopolylgalacturonase gene in the peach  
genome";  
RL Plant Cell Environ. 13:513-521(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Maravilla;  
RX MEDLINE=94302167; PubMed=8029352;  
RA Lester D.R., Speirs J., Orr G., Brady C.J.;  
RT "Peach (Prunus persica) endopolylgalacturonase - cDNA isolaton and mRNA  
analysis in melting and nonmelting peach cultivars";  
RL Plant Physiol. 105:225-231(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Maravilla;  
RA Speirs J.;

RL Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
(POLYLACTURONASES).  
DR EMBL: X77231; CAA54448.1; -.  
DR PIR: S71523; S71523.  
DR GO: GO:0005618; C:cell wall; IEA.  
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO: GO:0004650; F:polylacturonase activity; IEA.  
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro: IPR000743; Glyco\_hydro\_28.  
DR InterPro: IPR006526; Pbh1.  
DR InterPro: IPR00408; Reg\_chir\_condens.  
DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
DR SMART: SM00710; Pbh1; 6.  
DR PROSITE: PS00502; POLYLACTURONASE; 1.  
DR PROSITE: PS00626; RCCL\_2; 1.  
KW Cell wall, Glycosidase, Hydrolase.  
SQ SEQUENCE 458 AA, 49716 MW, E7BABBC8BB688A5 CRC64;  
Query Match 46.4%; Score 1108; DB 10; Length 458;  
Best Local Similarity 48.9%; Pred. No. 1.6e-64;  
Matches 226; Conservative 82; Mismatches 126; Indels 28; Gaps 8;  
QY 1 WTQNSILLIIIFASSISTCRSNVIDDNLFKQYVDNLEQEPFAHDFQAYLSTYS----- 56  
DB 1 MALQHLVLFYVVSFCAASCYSQFOEYNSLHSEVDH--EKEGNSRAHPSNMNTIEG 58  
QY 57 -----KNIESNNNDKV---DKNGIKVINLSFGAKGDKTYDNTAFEOANWAC 103  
DB 59 VKFMEFIFRAQLIESRKLIERAGSSSSSVKTIIVANFAKANGAD-DTRAEKAKKAC 117  
QY 104 SRTFVQFVFPKKNVYLLKQITFSGPCRSISVKTIFGSLEASSKISDYKORLMIAPDSV 163  
DB 118 SENGAIIVLVVPG-KYIVLRPIEFSGPKSHLTMQYGTIEASDDSVYDVTWMLFENV 176  
QY 164 QNLVVGCGGTINGNQVWMPSSCKINKSLPCGD-APTALTFWNCNKLKYNLKSNAQOI 222  
DB 177 QSLVAVGEGTINGNRWNENCKRKPQPCNEQAPLAVTFNKNLTVKNIKIDQAOQM 236  
QY 223 HIKFESCTNVVSNLMINASAKSPNTDGVHNSNTQYIQISDTTIGTDDCISIVSGSONV 282  
DB 237 HVRFQNCNKEASHLTVTPEDSPNTDGIHITNKITITSSSVITGDDCISIVSGSQRY 296  
QY 283 QATNITCGPHGISIGSLGSGNSSEAYVSNVTVNEAKIIGAENGVRKTMWGSGQASNIK 342  
DB 297 QATDITCGPHGISIGSLGEDNANDHVGVPFNGAKKISGTSNGVRKTMWGSGGSASNTV 356  
QY 343 FLNVEMQDYKPIIIDQNYCD-RVEPCIOQFSAVQKNNVYENIKGTSATKVAIKFDCST 401  
DB 357 FQNVEMNDVTNPIIIDQNYCDHKNDCTFQRSAAVQKNNLYONIRGTSASTDAITFNCQ 416  
QY 402 NPFCEGIIMNTLVGSEKPESEATCKNVHFNNAHEVTHC 443  
DB 417 SVPCQGIIVQNTQL-----QNAKACNNVKAIPKGAIVPRCS 453  
RESULT 12  
ID Q8VXT3 PRELIMINARY; PRT; 433 AA.  
AC Q8VXT3;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Endo polylacturonase.  
GN PG.  
OS Brassica rapa (Turnip).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Brassica.  
OC NCBI\_Taxid=51350;  
QY [1]  
RN SEQUENCE FROM N.A.  
RP Rodriguez-Gacio M., Nicolas C., Matilla A.;

```

RT "Isolation, characterization and expression of one
RL endopolygalacturonase from Brassica rapa."
CC Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases
CC -1 SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
(POLYGALACTURONASES).
DR EMBL; AJ428543; CAD21651.2; -.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0004650; F:polygalacturonase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR PROSITE; PS00502; POLYGALACTURONASE; 1.
DR PROSITE; PS00626; RCC1_2; 1.
DR Cell wall; Glycosidase; Hydrolase.
SQ SEQUENCE 433 AA; 46638 MW; BDD796980D1C653A CRC64;

Query Match 46.3%; Score 1106; DB 10; Length 433;
Best Local Similarity 50.0%; Pred. No. 2,1e-64;
Matches 222; Conservative 69; Mismatches 125; Indels 28; Gaps 7;

QY 7 SILLIIIFASSISTCSNVID-----DNLFKQYDNLBOEFADPOAYLSYSKNIES 61
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 9 AIFLCVILMLACCOLSSNVDDGYGHEDGSFQT--DSLILKLNDDVLTLSKSDRPTTES 66
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 62 NNNIDKVDKNGIKVINYLSFGAKGDKGTYPDNIAPFOAMNEACSSRTVPQFVPRKKNYLL 121
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 67 S-----TVSVSNFGAKGDKGTDPDQAFKKAMKACSTNGVTFLLIPGKTYLL 114
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 122 KQITFSGPCRSSISVKIFGSLKSSKISDY-KDRRLIADSVQNLVVGSG--GTINGNG 178
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 115 KSIRFRGPCKSLSPFOLLGTLASATKRSIDYNDKNHMLILEDVNNLSIDGSAQIVGNG 174
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 179 QVWMPSSCKINKSLPCDADPALTFTWCKLKNLKSNAQOQHIFESCTNVASNLML 238
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 175 KIWQNSCKIDKSKPCPKATALTLYNLKNLNVKNLEVRNAQOQIISIEKCNVDDVKNVK 234
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 239 INASAKSPNTDGVHVSNTQYIQISDTIIIGTDDCISIVSGSQNVQATNTGPGHGISIG 298
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 235 ITAPGDSFNTDGHIVATKIRISNSDITIGDDCISIEDSQNVQINDLTGCPHGISIG 294
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 299 SLGSGNSEAVSVNTVNEAKIIGAENGVRITKWGGSGQASNIKFLVEMQDVYPIIID 358
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 295 SLGDGNSKAVVSGINVDGATLSETDNGVRITKYGGSGGTAKNIKFQNIKRDNDVKNPIIID 354
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 359 QNYCDRYEPCIQOFSAVQVNNVYENIKGTSATKVAIKPDCSTNFPCEGIMENINLVE 418
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 355 QNYCDK-DKCEQOESAVQVNNVYRNIGTSAIDVAIMFNCVYKPCQGIIVLENNIKG- 412
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 419 SGKPSKATCNVHNNAEHTPHC 442
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 413 ----GRASCKNVNVKDKGTSPKC 432
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 13
Q42636 PRELIMINARY; PRT; 433 AA.
ID 042636
AC 042636;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Polygalacturonase precursor (EC 3.2.1.15).
OS Brassicaceae.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Brassica.
OX NCBI_Taxid=3708;
RN [1]
RP SEQUENCE FROM N.A.
RA Jenkins E.S., Paul W., Coupe S.A., Bell S., Davies L., Roberts J.A.;
RT "Characterization and expression of an mRNA encoding polygalacturonase

```

```

RT during pod development in oilseed rape (Brassica napus L.)";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases
CC -1 SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
(POLYGALACTURONASES).
DR EMBL; Z49971; CAA90272.1; -.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0004650; F:polygalacturonase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000743; Glyco_hydro_28.
DR InterPro; IPR006626; PBH1.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR SMART; SM00710; PBH1; 6.
DR PROSITE; PS00502; POLYGALACTURONASE; 1.
DR PROSITE; PS00626; RCC1_2; 1.
DR Cell wall; Glycosidase; Hydrolase; Signal.
FW SIGNAL.
FT CHAIN 1
SQ SEQUENCE 433 AA; 46670 MW; 39BE24426D26739 CRC64;

Query Match 46.2%; Score 1103; DB 10; Length 433;
Best Local Similarity 49.8%; Pred. No. 3,2e-64;
Matches 221; Conservative 70; Mismatches 125; Indels 28; Gaps 7;

QY 7 SILLIIIFASSISTCSNVID-----DNLFKQYDNLBOEFADPOAYLSYSKNIES 61
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 9 AIFLCVILMLACCOLSSNVDDGYGHEDGSFET--DSLILKLNDDVLTLSKSDRPTTES 66
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 62 NNNIDKVDKNGIKVINYLSFGAKGDKGTYPDNIAPFOAMNEACSSRTVPQFVPRKKNYLL 121
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 67 S-----TVSVSNFGAKGDKGTDPDQAFKKAMKACSTNGVTFLLIPGKTYLL 114
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 122 KQITFSGPCRSSISVKIFGSLKSSKISDY-KDRRLIADSVQNLVVGSG--GTINGNG 178
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 115 KSIRFRGPCKSLSPFOLLGTLASATKRSIDYNDKNHMLILEDVNNLSIDGSAQIVGNG 174
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 179 QVWMPSSCKINKSLPCDADPALTFTWCKLKNLKSNAQOQHIFESCTNVASNLML 238
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 175 KIWQNSCKIDKSKPCPKATALTLYNLKNLNVKNLEVRNAQOQIISIEKCNVDDVKNVK 234
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 239 INASAKSPNTDGVHVSNTQYIQISDTIIIGTDDCISIVSGSQNVQATNTGPGHGISIG 298
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 235 ITAPGDSFNTDGHIVATKIRISNSDITIGDDCISIEDSQNVQINDLTGCPHGISIG 294
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 299 SLGSGNSEAVSVNTVNEAKIIGAENGVRITKWGGSGQASNIKFLVEMQDVYPIIID 358
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 295 SLGDGNSKAVVSGINVDGATLSETDNGVRITKYGGSGGTAKNIKFQNIKRDNDVKNPIIID 354
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 359 QNYCDRYEPCIQOFSAVQVNNVYENIKGTSATKVAIKPDCSTNFPCEGIMENINLVE 418
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 355 QNYCDK-DKCEQOESAVQVNNVYRNIGTSAIDVAIMFNCVYKPCQGIIVLENNIKG- 412
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 419 SGKPSKATCNVHNNAEHTPHC 442
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 413 ----GRASCKNVNVKDKGTSPKC 432
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 14
Q8GTP8 PRELIMINARY; PRT; 460 AA.
ID 08GTP8
AC 08GTP8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Polygalacturonase precursor (EC 3.2.1.15).
OS Pyrus communis (Pear).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Rosales; Rosaceae; Maloideae; Pyrus.
OX NCBI_Taxid=23211;
RN [1]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. La France; TISSUE=fruit;  
 RA Hiwasa K., Kinugasa Y., Amano S., Hashimoto A., Nakano R., Inaba A.,  
 RA Kubo Y.;  
 RT "Ethylene is required for both the initiation and progression of  
 softening in pear (Pyrus communis L.) fruit.";  
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB084461; BAC268.1; -  
 DR GO; GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0004650; F:polygalacturonase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000743; Glyco\_Hydro\_28.  
 DR InterPro; IPR006626; Pbh1.  
 DR InterPro; IPR000408; Reg chr condens.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR SMART; SM00710; Pbh1; 5.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 DR PROSITE; PS00626; RCCL\_2; 1.  
 DR Signal; Hydrolyase; Glycosidase.  
 FT SIGNAL 1 24 POTENTIAL.  
 SQ SEQUENCE 460 AA; 49460 MW; 4F8A65AD1899591E CRC64;

Query Match 44.9%; Score 1073; DB 10; Length 460;  
 Best Local Similarity 47.9%; Pred. No. 3.2e-62; Indels 34; Gaps 11;  
 Matches 224; Conservative 78; Mismatches 132;

1 MVIQRNSILLITIFASISIST--CRSNVIDD-NLFRQVNDILDEOFADFCAYLSLS-56  
 1 MALKTOLMSFVVFVVFSTTSCSSSFOEVNALSVDHVDVDRSGNSRAVPSYMDI 60  
 57 -----KNIESNNNDKVDKNGI-----XVINVLSFGAKGDKTYDNIABE 96  
 61 IRLKXEMELIRPTQLSSSRKLNITI-TGIIATSSAPAKTISVDDFGAKKGAD-DTQAFV 118  
 97 QANNEKSSRTQFVPPKKNVILKQITSGPCRSISYKIFGSLASXSIDYDRRL 156  
 119 KAKKACSSSGAIVLVPPQ-KKYLVRPIDFSGPKCKQLVQIVGIIEASDRSIYKIDH 177  
 157 WIAFDSVQNLVVGAGGTINGQVWMPSSCKINKSLPC-ADAPALTFFMCKMLKVNLIK 215  
 178 WLFVDVQNLVVGPTINGSGNIMKNSCKRRKPPPCGTHAFTATYFNRCNNLVKNLK 237  
 216 SKNAQOIHIFESCTNVVNASNIMINASAKSPNTDGVHVSNTQYIQISDTTIGTDDCISI 275  
 238 IQDAQMHVAFQICINVOASRLVTVAPEDEPNTDGIHVNTQNTITSSSVIGTDDCISI 297  
 276 VSSGQVQATNTTCGFGHGISISLGSNSEAYVSNVTVEAKTIGAEKSVRIKTMQGS 335  
 298 VSSGQVQATDITCGFGHGISISLGEDSKDHVSGCVAGAKLSTGTSNGLRIKTMQGS 357  
 336 GQASNTKPLINVENQVYPIIDQNYCD-RVEPCIOQFSAVQVKNVYENIKGTSATKVA 394  
 358 GSAKNIVFQVQVNNVTNPPIIIDQNYCDHKTKCKQKQSAVQVKNVYQVIRGTSASGDA 417  
 QY 395 IKFDCSTNPECEGIIMENINLVGSGKPSSEATCKNVFNNAEHVTPHC 442  
 DB 418 ITLNCQSVPCRGIALQSVRL--QNGR---ACGNVQPAVKGVASPRC 460

## RESULT 15

065886 PRELIMINARY; PRT; 335 AA.  
 ID 065886;  
 AC 065886;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Polygalacturonase (Fragment).  
 GN RAS3.  
 OS Rubus idaeus (Raspberry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC eurosids I; Rosales; Rosaceae; Rosoideae; Rubus.  
 OC NCBI\_TaxID=32247;

RP [1]  
 RC STRAIN=CV. Glen Clova; TISSUE=fruit;  
 RA Jones C.S., Davies H.V., McNicol R.J., Taylor M.A.;  
 RT "Expression of three genes up-regulated in ripening raspberry fruit  
 (Rubus idaeus cv. Glen Clova)";  
 RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 (POLYGALACTURONASES).  
 DR EMBL; AJ224147; CA11846.1; -  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0004650; F:polygalacturonase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000743; Glyco\_hydro\_28.  
 DR InterPro; IPR006626; Pbh1.  
 DR InterPro; IPR000408; Reg chr condens.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR SMART; SM00710; Pbh1; 5.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 DR PROSITE; PS00626; RCCL\_2; 1.  
 DR Cell wall; Glycosidase; Hydrolyase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 335 AA; 36432 MW; B8BDFC7A85B6CDBA CRC64;

Query Match 44.4%; Score 1061.5; DB 10; Length 335;  
 Best Local Similarity 62.1%; Pred. No. 1.2e-61;  
 Matches 208; Conservative 45; Mismatches 71; Indels 11; Gaps 5;

112 VPKKNVYLLKQTFSGPCRSISYKIFGSLASXSIDY-KDRLMIAFDSVQNLVVG 170  
 7 LVPR-KNYLVKPTTFSGPKCKLTQIYGSIEDSDRSYVKDLYHMLFDVNRNLVQG 65  
 171 GGTINGQVWMPSSCKINKSLPCRD-APTALTFFMCKMLKVNLIKSKNAQOIHIFESC 229  
 66 PGTINGQVWMPSSCKINKSLPCRD-APTALTFFMCKMLKVNLIKSKNAQOIHIFESC 125  
 230 TNVVNASNIMINASAKSPNTDGVHVSNTQYIQISDTTIGTDDCISIVSGSQNTQATNITC 289  
 126 TNVQASVYLVTVAPEDEPNTDGIHVNTQNTITSSSVIGTDDCISIVSGSQNTQATNITC 185  
 290 GPHGISISLGSNSEAYVSNVTVEAKTIGAEKSVRIKTMQGSQASNIKFLINEMQ 349  
 186 GPHGISISLGSNSEAYVSNVTVEAKTIGAEKSVRIKTMQGSQASNIKFLINEMQ 245  
 350 DVKYPPIIDQNYCDRYE--PCIQQFSAVQVKNVYENIKGTSATKVAIKFDCSTNPECEG 407  
 246 DVNPIIIDQNYCDTSDKRCCKQKQSAVQVKNVYQVIRGTSASKVALAFDCSKSPICG 305  
 QY 408 IIMENINLVGSGKPSSEATCKNVFNNAEHVTPHC 442  
 DB 306 IYLVQVQL-----HRAKCSNVNLAVKENVSRC 334

Search completed: July 21, 2004, 17:08:12  
 Job time: 77.4797 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 21, 2004, 17:00:07 ; Search time 2.07709 Seconds

(without alignments)  
1360.306 Million cell updates/sec

Title: US-10-018-604-4

Sequence: 1 NNGANGQXV 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	96.0	10	2 AAR45389	Aar45389 PG beta s
2	48	96.0	10	2 AAW10696	Aaw10696 Tomato po
3	48	96.0	10	3 AAB19329	Aab19329 Polygalac
4	48	96.0	10	4 AAB48353	Aab48353 Polygalac
5	40	80.0	979	5 ABB35664	Abp35664 Fungal ZB
6	39	78.0	321	4 AAB70839	Abb70839 S. halste
7	39	78.0	449	2 AAW18210	Aaw18210 Cellulom
8	38	76.0	427	4 ABB61028	Abb61028 Drosophil
9	37	74.0	149	3 AAY44604	Aay44604 Maize MLO
10	37	74.0	443	4 ABB61160	Abb61160 Drosophil
11	37	74.0	568	6 ADA07853	Ada07853 pinctada
12	37	74.0	582	4 ABB59482	Abb59482 Drosophil
13	37	74.0	1038	6 ABB15903	Abb15903 Protein e
14	36	72.0	141	5 ABB83327	Abb83327 Murine TS
15	36	72.0	175	6 ABB852643	Abb852643 Protein s
16	36	72.0	220	4 ABB71175	Abb71175 Drosophil
17	36	72.0	296	4 ABB65333	Abb65333 Drosophil
18	36	72.0	420	5 ABB73816	Abb73816 Candida a
19	36	72.0	887	6 ABB21462	Abb21462 Protein e
20	36	72.0	1018	4 AAB37245	Aab37245 Steaphyloc
21	36	72.0	1018	4 AAB34301	Aab34301 Steaphyloc
22	36	72.0	1018	6 ABB18922	Abb18922 Pathogen
23	36	72.0	1284	1 AAB91187	Aap91187 Sequence
24	36	72.0	2248	4 ABB64494	Abb64494 Drosophil
25	35	70.0	10	5 ABB76722	Abb76722 Peptide 1

26	35	70.0	19	2 AAB60371	Aab60371 Vasopress
27	35	70.0	86	5 ABB34678	Abb34678 Human ORF
28	35	70.0	121	2 AAY17484	Aay17484 Human Nix
29	35	70.0	125	6 ABB25282	Abb25282 Mouse BAC
30	35	70.0	202	4 ABB61097	Abb61097 Mouse BAC
31	35	70.0	206	5 ABB83329	Abb83329 Murine TS
32	35	70.0	208	6 ABB70357	Abb70357 Human adl
33	35	70.0	218	2 AAY17483	Aay17483 Murine N1
34	35	70.0	219	2 AAB93864	Aab93864 Human BNI
35	35	70.0	219	5 ABB83328	Abb83328 Human TSA
36	35	70.0	219	5 ABB65128	Abb65128 Hypoxia-r
37	35	70.0	232	2 AAB63114	Aab63114 A human a
38	35	70.0	246	4 ABB64176	Abb64176 Drosophil
39	35	70.0	328	5 ABB73597	Abb73597 Candida a
40	35	70.0	411	6 ABB07280	Abb07280 Allotococ
41	35	70.0	424	6 ABB07282	Abb07282 Allotococ
42	35	70.0	455	4 AAB81952	Aab81952 Pea bligh
43	35	70.0	463	4 ABB71163	Abb71163 Drosophil
44	35	70.0	495	5 ABB93381	Abb93381 C. albica
45	35	70.0	525	6 ABB36839	Abb36839 Protein e

## ALIGNMENTS

RESULT 1	
AAR45389	
ID	AAR45389 standard; protein; 10 AA.
XX	
AC	AAR45389;
XX	
DT	25-MAR-2003 (revised)
DT	06-JUL-1994 (first entry)
XX	
DE	PG beta subunit internal fragment.
XX	
KW	Beta subunit; PG; polygalacturonase; tomato; pathogen.
OS	Lycopersicon esculentum.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 9 /note="unknown"
XX	
PN	BP577252-A1.
XX	
PD	05-JAN-1994.
XX	
PF	07-MAY-1993; 93EP-00303533.
XX	
PR	08-MAY-1992; 92US-00880915.
XX	
PA	(UYAR-) UNIV ARIZONA STATE.
XX	
PI	Delilepenna D;
DR	WPI; 1994-009528/02.
XX	
PT	Transgenic tomato plants contg. lowered levels of polygalacturonase
PT	isoform 1 - for prodn. of tomatoes less susceptible to pathogens and
PT	requiring less energy to inactivate poly-galacturonase activity.
XX	
PS	Example; Page 6; 15pp; English.
XX	
CC	N-terminal sequence analysis of purified PG beta-subunit was performed.
CC	Two internal fragments (AAR45389-90) were used for the construction of
CC	primers (AAQ54493-95). Two beta-subunit cDNA clones and fragments of 11
CC	others were sequenced. The consensus DNA sequence obtained is given in
CC	AAQ54492. (Updated on 25-MAR-2003 to correct PN field.)
XX	
SQ	Sequence 10 AA;
Query Match	96.0%; Score 48; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
| | | | |  
Db 1 NNGANGQXV 10

## RESULT 2

AAW10696  
ID AAW10696 standard; peptide; 10 AA.

XX AAW10696;

XX 25-MAR-2003 (revised)

DT 09-APR-1997 (first entry)

XX Tomato polygalacturonase isoform 1 beta subunit internal peptide.

XX Tomato; high soluble pectin content; recombinant; antisense; low cost;  
KM transgenic; beta subunit.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FT Misc-difference 9 /note= "unspecified amino acid"

XX US569831-A.

XX 29-OCT-1996.

XX 11-JUL-1994; 94US-00273538.

XX 08-MAY-1992; 92US-00880915.

XX (UYAR-) UNIV ARIZONA.

XX Dellapenna D;

XX WPI; 1996-496969/49.

XX Transgenic tomato plants with low level of polygalacturonase isoform 1 -  
PT have higher content of soluble pectin(s) and require less heat energy to  
PT eliminate polygalacturonase activity of the fruit.

XX Example 2; Col 15-16; 19pp; English.

XX AAW10696 and AAW10697 are internal peptides of the beta subunit of tomato  
CC polygalacturonase isoform 1 (PG1). The two peptides were generated by  
CC digestion with Lys-C and Gly-C endoproteases respectively. The peptides  
CC were used for the design of degenerate PCR primers used for cDNA library  
CC screening. A cDNA molecule encoding PG1 was identified, this sequence was  
CC used to produce constructs, antisense to the PG1 coding sequence, which  
CC are used to transform a tomato plant cell so as to reduce or eliminate  
CC PG1 activity. There are two isoforms of tomato polygalacturonase (PG),  
CC PG1 and PG2. PG1 has a much higher heat stability level (usually at least  
CC by 20 deg. C.) than PG2. By reducing or eliminating PG1 expression,  
CC leaving the less heat-tolerant isoform PG2. The process for production of  
CC tomato pastes, sauces etc. becomes less costly due to a lower temperature  
CC being needed for PG breakdown. A higher viscosity is also produced due to  
CC increased soluble pectin content. (Updated on 25-MAR-2003 to correct PF  
CC field.)

XX Sequence 10 AA;

QY Query Match 96.0%; Score 48; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
| | | | |  
Db 1 NNGANGQXV 10

## RESULT 3

AAB19329  
ID AAB19329 standard; peptide; 10 AA.

XX AAB19329;

XX 19-FEB-2001 (first entry)

XX Polygalacturonase beta-subunit peptide produced after Lys-C digestion.

XX Polygalacturonase beta-subunit; promoter; tomato; fruit; gene expression;  
XX transgenic plant.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

FT Misc-difference 9 /note= "unknown amino acid"

XX US6127179-A.

XX 03-OCT-2000.

XX 17-APR-1996; 96US-00632806.

XX 17-APR-1996; 96US-00632806.

XX (DELL/) DELLAPENNA D.

XX (WATS/) WATSON C F.

XX Dellapenna D, Watson CF;

XX WPI; 2000-611059/58.

XX Tomato polygalacturonase beta-subunit promoter, useful for controlling  
PT expression of target genes which affect the quality of developing fruit  
PT without affecting the ripening process.

XX Example; Col 9; 15pp; English.

XX The present sequence represents an internal fragment of the tomato  
CC polygalacturonase beta-subunit. The specification describes a  
CC polygalacturonase beta-subunit promoter. The promoter exhibits both high  
CC expression and high specificity, and provides tissue-preferential and  
CC developmentally regulated expression. The beta-subunit promoter is highly  
CC active in developing tomato fruit from as early as 10 days after  
CC pollination upto the onset of ripening, at which time it becomes  
CC inactive. The fruit specific, developmentally regulated promoter, can  
CC control the expression of target genes which affect the quality of  
CC developing fruit without affecting the ripening process. Target genes  
CC include genes responsible for sugar or starch metabolism, source sink  
CC relation, organic acid balance, flavour components, pathogen resistance,  
CC soluble solids, and water/pH relations. All or part of the promoter may  
CC be used to direct the expression of foreign or endogenous genes in the  
CC sense orientation, or in the antisense orientation, for overexpression or  
CC cosuppression

XX Sequence 10 AA;

QY Query Match 96.0%; Score 48; DB 3; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
| | | | |  
Db 1 NNGANGQXV 10

## RESULT 4

AAB48353  
ID AAB48353 standard; peptide; 10 AA.



```

XX AAB48353;
AC
XX
XX
XX 20-APR-2001 (first entry)
DT
XX
XX Polyalacturonase (PG) enzyme internal Lys-C peptide fragment.
DE
XX
XX Pectin; pectin methyltransferase; PME; polygalacturonase; PG; tomato; PTOM6;
KW food product; yogurt; milk; fruit juice; whey drink; de-esterification.
XX
XX Lycopersicon esculentum.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 9
FT /note="unknown"
FT
XX
XX WO20078982-A1.
PN
XX
XX 28-DEC-2000.
PD
XX
XX 15-JUN-2000; 2006WO-IB000869.
PF
XX
XX 17-JUN-1999; 99GB-00014209.
PR
XX
XX (DANI-) DANISCO AS.
PA
XX
XX Christensen TMB, Kreiberg JD;
PI
XX
XX WPI; 2001-091573/10.
DR
XX
XX
XX Modifying pectin, for foodstuffs preparation, involves transforming host
PT having pectin methyltransferase (PME) and polygalacturonase (PG) activity by
PT silencing PG activity, to increase PME to PG ratio.
XX
XX Example 2, Page 42; 78pp; English.
XX
XX The invention provides a new method for modifying pectin that involves
CC providing a host having pectin methyltransferase (PME) activity and
CC polygalacturonase (PG) activity, transforming the host by silencing PG
CC activity to provide an increased PME to PG ratio, preparing a PME extract
CC from the transformed host, and using the PME extract to modify pectin. A
CC PME modified pectin is useful for foodstuffs preparation, and to impart
CC an increased functionality to food products such as yogurt, milk/fruit
CC juice and whey drinks. PME is useful to reduce the number of ester groups
CC in a pectin in a block-wise manner, and to de-esterify two or more
CC adjacent galacturonic acid residue of a pectin on at least substantially
CC all of the pectin chains. The present sequence represents an internal
CC peptide fragment of the PG enzyme, used for designing primers
CC
XX
XX
SQ Sequence 10 AA;

```

```

Query Match          96.0%; Score 48; DB 4; Length 10;
Best Local Similarity 100.0%; Pred.No.0.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NNGANGGXV 10
   |||||
   |||||
   |||||
   |||||
   |||||
Db 1 NNGANGGXV 10

```

```

RESULT 5
AAB35664
ID AAB35664 standard; protein; 979 AA.
XX
XX AAB35664;
AC
XX
XX 24-JUL-2002 (first entry)
DT
XX
XX Fungal ZBC protein sequence #90.
DE
XX
XX Secondary metabolite; fungus; zinc binuclear cluster protein;
KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;

```

```

KW angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
KW insecticide; antineoplastic.
XX
XX Unidentified.
XX
XX WO200224865-A2.
PN
XX
XX 28-MAR-2002.
PD
XX
XX 19-SEP-2001; 2001WO-US029288.
PF
XX
XX 19-SEP-2000; 2000US-0233564P.
PR
XX
XX (MICR-) MICROBIA INC.
PA
XX
XX Holtzman D, Madden K, Maxon M, Sherman A;
PI
XX
XX WPI; 2002-352005/38.
DR
XX
XX N-PSDB; AEN79853.
DR
XX
XX
XX New method for improving the production of a secondary metabolite e.g.
PT antineoplastic agent, ergot alkaloid from a fungus involves modulation of
PT the expression of at least one zinc binuclear cluster protein gene.
XX
XX Disclosure; SEQ ID NO 192; 49pp + Sequence Listing; English.
XX
XX The invention relates to improving the production of a secondary
CC metabolite by a fungus. This involves modulating the expression of at
CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
CC improve the yield of the secondary metabolite. Methods of the invention
CC may be used for improving the production of the secondary metabolite e.g.
CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),
CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such
CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,
CC a fungal toxin, a modulator of cell surface receptor signalling, a plant
CC growth regulator, a pigment, an insecticide, or an antineoplastic
CC compound. The method results in a decrease in fermentor run-time, a
CC decrease in the size of the fermentor required for the production of
CC equivalent amounts of the secondary metabolite, or a decrease in the
CC biomass required for the production, which translates into decreased
CC waste that must be handled in downstream processing. The sequences given
CC in records AAB35575-AAB35722 represent ZBC proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained directly from WIPO at
CC ftp.wipo.int/pub/published_pot_sequences
CC
XX
XX
SQ Sequence 979 AA;

```

```

Query Match          80.0%; Score 40; DB 5; Length 979;
Best Local Similarity 100.0%; Pred.No.2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NNGANG 7
   |||||
   |||||
   |||||
   |||||
   |||||
Db 190 NNGANG 196

```

```

RESULT 6
AAB70839
ID AAB70839 standard; protein; 321 AA.
XX
XX AAB70839;
AC
XX
XX 06-ANG-2003 (revised)
DT
XX
XX 25-JUN-2001 (first entry)
DT
XX
XX S. halstedii cellulase protein.
DE
XX
XX Cellulase; fusion construct; protein synthesis; papermaking; medicine;
KW glyceralddehyde phosphate dehydrogenase; food industry; feed industry;
KW washing composition; catalyst.

```

```

XX OS Streptomyces halstedii.
XX XX
XX PN DE19946047-A1.
XX XX
XX XX 29-MAR-2001.
XX XX
XX PF 25-SEP-1999; 99DE-01046047.
XX PR 25-SEP-1999; 99DE-01046047.
XX XX
XX PA (BADI ) BASF AG.
XX PI Althoefer H, Pompejus M, Revuelta Doval JL, Santos M, Jimenez A,
XX PI Benito R, Santamaria R, Fernandez J,
XX XX
XX DR WPI: 2001-291995/31.
XX DR N-PSDB; AAF61508.
XX XX
XX PT Modified Ashbya organisms with altered protein production properties,
XX PT useful particularly for production of Streptomyces halstedii cellulase.
XX XX
XX PS Example 1; Fig 1; 12pp; German.
XX XX
XX CC This invention describes a novel microorganism (A) of the genus Ashbya
XX CC for biotechnical production of proteins (I), containing a genetic
XX CC construct so that the protein-synthesis capacity is altered relative to
XX CC the wild-type species A. gossypii ATCC 10895. The invention also
XX CC describes (1) cellulase gene (Iia), for use in (A), having a sequence,
XX CC from Streptomyces halstedii; (2) promoter (P) and transcriptional
XX CC terminator (T) from the sequence encoding the glyceraldehyde phosphate
XX CC dehydrogenase of A. gossypii; (3) genetic structure containing (Iia) plus
XX CC regulatory regions that provide expression in Ashbya; (4) a vector
XX CC containing (Iia), P and/or T, or the structure of (3), plus signals that
XX CC provide replication in host cells or integration into a cellular genome;
XX CC (5) transformed Ashbya, for production of cellulase, containing the
XX CC structure of (3) in replicable form; (6) producing (1) using (A); (7)
XX CC producing (A) by exchanging the promoter and/or increasing the gene copy
XX CC number, and/or by introducing a cellulase-encoding gene; and (8) (I)
XX CC produced by the method of (6). (A) are used for recombinant production of
XX CC endogenous or exogenous proteins, specifically cellulase, useful in
XX CC papermaking, medicine, food/feed industries, washing compositions and/or
XX CC as catalysts. (A) provide higher protein yields than wild-type cells.
XX CC This sequence represents the S. halstedii cellulase protein which is used
XX CC in the construction of the A. gossypii/S. halstedii genetic construct
XX CC used to illustrate the method of the invention. (Updated on 06-AUG-2003
XX CC to correct OS field.)
XX XX
XX SQ Sequence 321 AA;
XX XX
XX QY Query Match 78.0%; Score 39; DB 4; Length 321;
XX DB Best Local Similarity 75.0%; Pred. No. 1.2e+02;
XX Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX 1 NGNGANGQ 8
XX 256 NGNGSNGE 263
XX XX
XX RESULT 7
XX ID AAM18210 standard; protein; 449 AA.
XX XX
XX AC AAM18210;
XX XX
XX DT 25-MAR-2003 (revised)
XX DT 29-SEP-1997 (first entry)
XX XX
XX DE Cellulomonas fimi endoglucanase.
XX XX
XX KM Endoglucanase; cema gene; cellulase; cellulose; glucose.
XX OS Cellulomonas fimi.
XX XX

```

```

XX XX Key Location/Qualifiers
XX FH Peptide 1..31
XX FT /label= Sig_peptide
XX FT Protein 32..449
XX FT /label= Mat_protein
XX FT Region 143..165
XX FT /note= "Pro-Thr-rich conserved sequence of unknown
XX FT function"
XX XX
XX PN US5643791-A.
XX XX
XX PD 01-JUL-1997.
XX XX
XX PF 19-JAN-1994; 94US-00185303.
XX XX
XX PR 07-AUG-1986; 86US-00894326.
XX PR 18-DEC-1990; 90US-00630396.
XX PR 19-JUN-1992; 92US-00902460.
XX XX
XX PA (UYBR-) UNIV BRITISH COLUMBIA.
XX XX
XX PI Kilburn DG, Miller RC, Warren RAJ;
XX XX
XX DR WPI: 1997-350249/32.
XX DR N-PSDB; AAT71711.
XX XX
XX FT Cloned Cellulomonas fimi cema gene - encodes endo:glucanase enzyme, for
XX FT conversion of cellulose to glucose.
XX XX
XX PS Disclosure; Fig 5A-F; 20pp; English.
XX XX
XX CC The amino acid sequence (AAM18210) of a Cellulomonas fimi endo-1,4- beta-
XX CC glucanase (EC-3.2.1.4) was deduced from an isolated DNA fragment
XX CC (AAT71711) contg. the cema gene. The enzyme is capable of hydrolysing
XX CC beta-1,4-glycosidic linkages in the interior of a cellulose molecule,
XX CC converting the cellulose to glucose. The C-terminal region of the enzyme
XX CC is crucial for activity; deletion of 12 amino acids from that end results
XX CC in loss of all activity. Recombinant endoglucanase can be expressed in
XX CC host cells, esp. E. coli, utilising vectors incorporating the cema gene.
XX CC (Updated on 25-MAR-2003 to correct PF field.)
XX XX
XX SQ Sequence 449 AA;
XX XX
XX QY Query Match 78.0%; Score 39; DB 2; Length 449;
XX DB Best Local Similarity 75.0%; Pred. No. 1.7e+02;
XX Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX 1 NGNGANGQ 8
XX 381 NGNGSNGE 388
XX XX
XX RESULT 8
XX ID ABB61028 standard; protein; 427 AA.
XX XX
XX AC ABB61028;
XX XX
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Drosophila melanogaster polypeptide seq ID NO 9876.
XX XX
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX XX
XX OS Drosophila melanogaster.
XX XX
XX PN W0200171042-A2.
XX XX
XX PD 27-SEP-2001.
XX PD 23-MAR-2001; 2001WO-US009231.
XX PF

```

```

XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR MPI; 2001-656860/75.
DR N-PSDB; ABL05131.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 9876; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB172072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 427 AA:
XX
Query Match 76.0%; Score 38; DB 4; Length 427;
Best Local Similarity 70.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NNGANGQXY 10
DB 250 HNGANGNHV 259

```

```

PI Briggs SP, Simmons CR;
XX
XX MPI; 2000-137196/12.
DR N-PSDB; AA249563.
XX
PT Creating or enhancing disease resistance in plants by modulating plant
PT resistance genes expression.
XX
PS Claim 1b; Page 70; 88pp; English.
XX
CC The present sequence is the maize MLO5 (ZmMLO5) protein, which exhibits
CC sequence homology to barley MLO1 (HvMLO1) sequence. MLO5 protein is
CC encoded by a mutation-induced recessive allele MLO5, located on
CC chromosome 3, that confers resistance to plant pathogens. Expression of
CC native MLO genes in plants can be altered by transforming them with a DNA
CC construct comprising the mutated Mlo gene. Decreasing the expression or
CC activity of native MLO protein leads to enhanced resistance of plants
CC against pathogens such as fungi, virus, nematodes and insects. Mlo gene
CC can be used to isolate related sequences from other plants and as
CC molecular markers used in breeding programs aimed at improving disease
CC resistance. MLO protein has anti-infective, antipathogenic, antiviral,
CC antifungal, antihelminthic and anti-arthropod activity
XX
SQ Sequence 149 AA:
XX
Query Match 74.0%; Score 37; DB 3; Length 149;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNGANG 7
DB 96 NNGSNG 102

```

```

RESULT 9
AA44604
ID AA44604 standard; protein; 149 AA.
XX
AC AA44604;
XX
DT 07-APR-2000 (first entry)
XX
DE Maize MLO5 protein.
XX
KM MLO5 protein; ZmMLO5; chromosome 3; mutation; recessive allele;
KM disease resistance; pathogen; anti-infective; antipathogenic; antiviral;
KM antifungal; antihelminthic; anti-arthropod; maize; HvMLO1.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT Region 1..37
FT Domain 38..61
FT Region /label= Intervening_region
FT Region /label= Transmembrane_helix
FT Region 62..149
FT Region /label= C-terminal_region
XX
PN MO200001722-A1.
XX
PD 13-JAN-2000.
XX
PF 07-JUL-1999; 99MO-US015255.
XX
PR 07-JUL-1998; 98US-0091875P.
XX
PA (PION-) PIONEER HT-BRED INT INC.
XX

```

```

RESULT 10
ABB61160
ID ABB61160 standard; protein; 443 AA.
XX
AC ABB61160;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10272.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR MPI; 2001-656860/75.
DR N-PSDB; ABL05263.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 10272; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

```

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS7737-  
CC ABBS72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 443 AA;

Query Match 74.0%; Score 37; DB 4; Length 443;  
Best Local Similarity 75.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 8  
Db 111 NGNGANG 118

## RESULT 11

ADA07853  
ID ADA07853 standard; protein; 568 AA.

XX AC ADA07853;

XX DT 06-NOV-2003 (first entry)

XX DE Pinctada maxima nacrein.

XX KW nacrein; nacrein; pearl oyster cultivation; pearl formation.

XX OS Pinctada maxima.

XX PN US2003027258-A1.

XX PD 06-FEB-2003.

XX PF 27-FEB-2002; 2002US-00086510.

XX PR 02-AUG-2001; 2001US-0310070P.

XX PA (CHAN/) CHANG F F.

XX PA (LIH/) LI H.

XX PA (HSIE/) HSIEH-LI H M.

XX PI Chang FF, Li H, Hsieh-Li HM;

XX DR WPI; 2003-596340/56.

XX DR N-PSDB; ADA07856.

XX PT New isolated polynucleotide and polypeptide from the nacre gene of the  
XX Pinctada margaritifera species, useful for pearl oyster cultivation.

XX PS Disclosure; Page 24-25; 29pp; English.

XX CC The invention relates to an isolated polynucleotide that encodes nacrein.  
XX CC The methods and compositions of the present invention are useful for  
XX CC pearl oyster cultivation and pearl formation. The present sequence  
XX CC represents the amino acid sequence of Pinctada maxima nacrein.

XX SQ Sequence 568 AA;

Query Match 74.0%; Score 37; DB 6; Length 568;  
Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNGANG 7  
Db 387 NGNGANG 393

## RESULT 12

ABBS9482  
ID ABBS9482 standard; protein; 582 AA.

XX AC ABBS9482;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5238.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABU03585.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions.

XX PS Disclosure; SEQ ID NO 5238; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
XX CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS7737-  
XX CC ABBS72072). The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 582 AA;

Query Match 74.0%; Score 37; DB 4; Length 582;  
Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNGANG 7  
Db 440 NGNGANG 446

## RESULT 13

ABU15903  
ID ABU15903 standard; protein; 1038 AA.

XX AC ABU15903;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by prokaryotic essential gene #1430.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus aureus.

XX PN WO200277183-A2.



PI Marzioch M, Schultz JD, Supertl-Furga GD;  
XX  
DR WPI; 2003-250078/25.  
DR N-PSDB; ACC60685.

XX New isolated protein complexes useful for diagnosing a disease or  
PT disorder, or as a target for an active agent of a pharmaceutical,  
PT preferably a drug target in the treatment or prevention of disease or  
PT disorder.

XX  
PS Disclosure; SEQ ID NO 151; 17pp + Sequence Listing; English.

CC The invention relates to multiprotein complexes from eukaryotes. Proteins  
CC of the invention and DNA sequences encoding them are given in records  
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are  
CC obtainable by using a protein as a bait and isolating the set of proteins  
CC which is attached thereto from cells. Such protein complexes may comprise  
CC up to 30 distinct proteins. Protein complexes of the invention are useful  
CC for diagnosing a disease or disorder, or as a target for an active agent  
CC of a pharmaceutical, preferably a drug target in the treatment or  
CC prevention of a disease or disorder. Note: The sequence data for this  
CC patent is not represented in the printed specification, but is based on  
CC sequence information supplied by the European Patent Office. The complete  
CC document is available on CD-ROM

XX  
SQ Sequence 175 AA;

Query Match 72.0%; Score 36; DB 6; Length 175;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7  
Db 62 NGNGANG 68

Search completed: July 21, 2004, 17:06:23  
Job time : 5.07709 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2004, 17:04:52 ; Search time 0.620985 Seconds  
(without alignments)  
831.356 Million cell updates/sec

Title: US-10-018-604-4  
Perfect score: 50  
Sequence: 1 NNGANGQXV 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCUTS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/Backfillset.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	96.0	10	1	US-08-273-538A-1
2	48	96.0	10	3	US-08-632-806A-1
3	39	78.0	286	3	US-09-286-691-27
4	39	78.0	286	3	US-09-687-147-27
5	37	74.0	149	4	US-09-347-650-8
6	35	70.0	19	2	US-08-811-492-127
7	35	70.0	232	2	US-08-773-910-1
8	35	70.0	232	4	US-09-199-892-1
9	35	70.0	455	4	US-09-240-639-10
10	35	70.0	574	4	US-09-142-623-13
11	35	70.0	613	4	US-09-508-264A-3
12	35	70.0	1104	4	US-09-268-347-28
13	35	70.0	1104	4	US-09-268-347-34
14	34	68.0	11	6	5178861-8
15	34	68.0	50	1	US-08-178-477B-16
16	34	68.0	157	4	US-09-025-769B-372
17	34	68.0	157	4	US-09-025-769B-373
18	34	68.0	159	3	US-09-027-449-53
19	34	68.0	159	3	US-08-804-444A-53
20	34	68.0	159	3	US-09-026-985-53
21	34	68.0	159	4	US-09-121-952A-53
22	34	68.0	159	4	US-09-234-340A-53
23	34	68.0	163	4	US-09-025-769B-282
24	34	68.0	211	1	US-08-276-852-34
25	34	68.0	211	1	US-08-133-011-16
26	34	68.0	211	1	US-08-322-730A-16
27	34	68.0	211	1	US-08-367-874-16

28	34	68.0	211	1	US-08-899-575-34	Sequence 34, Appl
29	34	68.0	211	1	US-08-899-575-34	Sequence 34, Appl
30	34	68.0	211	2	US-08-383-619-16	Sequence 16, Appl
31	34	68.0	211	3	US-08-907-739-16	Sequence 16, Appl
32	34	68.0	211	3	US-09-729-597-16	Sequence 16, Appl
33	34	68.0	211	5	PCT-US93-08364-16	Sequence 16, Appl
34	34	68.0	211	5	PCT-US95-08743-34	Sequence 16, Appl
35	34	68.0	232	2	US-08-829-876-103	Sequence 103, App
36	34	68.0	232	4	US-09-234-873A-103	Sequence 103, App
37	34	68.0	232	4	US-09-234-873A-103	Sequence 103, App
38	34	68.0	233	2	US-08-829-876-101	Sequence 101, App
39	34	68.0	233	2	US-08-829-876-105	Sequence 105, App
40	34	68.0	233	3	US-08-438-745-15	Sequence 15, Appl
41	34	68.0	233	3	US-08-438-745-15	Sequence 17, Appl
42	34	68.0	233	3	US-09-219-019-15	Sequence 15, Appl
43	34	68.0	233	3	US-09-219-019-17	Sequence 17, Appl
44	34	68.0	233	4	US-09-234-874A-101	Sequence 101, App
45	34	68.0	233	4	US-09-234-874A-105	Sequence 105, App

#### ALIGNMENTS

RESULT 1  
US-08-273-538A-1  
; Sequence 1, Application US/08273538A  
; Patent No. 5569831  
; GENERAL INFORMATION:  
; APPLICANT: Dellapenna, Dean  
; TITLE OF INVENTION: Transgenic Tomato Plants with  
; TITLE OF INVENTION: Altered Polygalacturonase Isoforms  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Nicholas J. Seay  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/273, 538A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608/251-5000  
; TELEFAX: 608/251-9166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
; ORGANISM: Lycopersicon esculentum  
; US-08-273-538A-1

Query Match 96.0%; Score 48; DB 1; Length 10;  
Best local Similarity 100.0%; Pred. No. 0.041;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
|||||||

Db 1 NNGANGQXV 10

RESULT 2

US-08-632-806A-1

Sequence 1, Application US/08632806A

Patent No. 6127179

GENERAL INFORMATION:

APPLICANT: Dellapenna, Dean

APPLICANT: Watson, Colin P.

TITLE OF INVENTION: Gene Promoter for Tomato Fruit

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/632,806A

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 920214.90123

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9186

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Lycopersicon esculentum

US-08-632-806A-1

Query Match 96.0%; Score 48; DB 3; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10

Db 1 NNGANGQXV 10

RESULT 3

US-09-286-691-27

Sequence 27, Application US/09286691

Patent No. 6190189

GENERAL INFORMATION:

APPLICANT: Li, Xin-liang

APPLICANT: Ljungdahl, Lars G.

APPLICANT: Chen, Huizhong

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96

CURRENT APPLICATION NUMBER: US/09/286,691

CURRENT FILING DATE: 1999-04-05

EARLIER APPLICATION NUMBER: US 60/027,883

EARLIER FILING DATE: 1996-10-04

EARLIER APPLICATION NUMBER: PCT US97/18008

EARLIER FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 27

LENGTH: 286

TYPE: PRT

ORGANISM: Cellulomonas fimi

US-09-286-691-27

Query Match 78.0%; Score 39; DB 3; Length 286;

Best Local Similarity 75.0%; Pred. No. 38;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQ 8

Db 218 NNGSNGE 225

RESULT 4

US-09-687-147-27

Sequence 27, Application US/09687147

Patent No. 6268198

GENERAL INFORMATION:

APPLICANT: Li, Xin-liang

APPLICANT: Ljungdahl, Lars G.

APPLICANT: Chen, Huizhong

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96A

CURRENT APPLICATION NUMBER: US/09/687,147

CURRENT FILING DATE: 2000-10-12

PRIOR APPLICATION NUMBER: US 60/027,883

PRIOR FILING DATE: 1996-10-04

PRIOR APPLICATION NUMBER: PCT US97/18008

PRIOR FILING DATE: 1997-10-03

PRIOR APPLICATION NUMBER: 09/286,691

PRIOR FILING DATE: 1999-04-05

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 27

LENGTH: 286

TYPE: PRT

ORGANISM: Cellulomonas fimi

US-09-687-147-27

Query Match 78.0%; Score 39; DB 3; Length 286;

Best Local Similarity 75.0%; Pred. No. 38;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQ 8

Db 218 NNGSNGE 225

RESULT 5

US-09-347-650-8

Sequence 8, Application US/09347650

Patent No. 6576814

GENERAL INFORMATION:

APPLICANT: Briggs, Steven

TITLE OF INVENTION: Manipulation of M10 Genes to Enhance Disease Resistance

FILE REFERENCE: 5718-42035718/158714

CURRENT APPLICATION NUMBER: US/09/347,650

CURRENT FILING DATE: 1999-07-06

NUMBER OF SEQ ID NOS: 20

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 8

LENGTH: 149

TYPE: PRT

ORGANISM: Zea mays

FEATURE:

OTHER INFORMATION: M105

US-09-347-650-8

Query Match 74.0%; Score 37; DB 4; Length 149;



Best Local Similarity 85.7%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
|||||  
DB 96 NNGANG 102

RESULT 6  
US-08-811-492-127  
; Sequence 127, Application US/08811492  
; Patent No. 5834247  
; GENERAL INFORMATION:  
; APPLICANT: COMB, DONALD G.  
; APPLICANT: PERLER, FRANCINE B.  
; APPLICANT: JACK, WILLIAM E.  
; APPLICANT: XU, MING-QUN  
; APPLICANT: HODGES, ROBERT A.  
; APPLICANT: NOREN, CHRISTOPHER J.  
; APPLICANT: CHONG, SHARONG S.C.  
; APPLICANT: ADAM, ERIC  
; APPLICANT: SOUTHWORTH, MAURICE  
; TITLE OF INVENTION: MODIFIED PROTEINS, METHODS OF THEIR  
; TITLE OF INVENTION: PRODUCTION AND METHODS FOR PURIFICATION OF TARGET  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 155  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.  
; STREET: 32 TOZER ROAD  
; CITY: BEVERLY  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01915  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC\DOS\MS\ DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,492  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/580,555  
; FILING DATE: 29-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/496,247  
; FILING DATE: 28-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/146,885  
; FILING DATE: 03-NOV-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/004,139  
; FILING DATE: 09-DEC-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Gregory D  
; REGISTRATION NUMBER: 30901  
; REFERENCE/DOCKET NUMBER: NEB-036C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-927-5054  
; TELEFAX: 509-927-1705  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 127:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide

US-08-811-492-127

Query Match 70.0%; Score 35; DB 2; Length 19;  
Best Local Similarity 85.7%; Pred. No. 9.8;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANG 7  
|||||  
DB 4 NNGANG 10

RESULT 7  
US-08-773-910-1  
; Sequence 1, Application US/08773910  
; Patent No. 5847093  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hallman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HUMAN APOPTOSIS REG  
; TITLE OF INVENTION: ULATOR  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/773,910  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0184 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 232 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 1715374  
; US-08-773-910-1  
Query Match 70.0%; Score 35; DB 2; Length 232;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANG 7  
|||||  
DB 49 NNGANG 55

RESULT 8  
US-09-199-892-1  
; Sequence 1, Application US/09199892  
; Patent No. 6307021

GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Suya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL HUMAN APOPTOSIS REG  
TITLE OF INVENTION: ULATOR  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/199,892  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0184 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 232 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1715374  
US-09-199-892-1

Query Match 70.0%; Score 35; DB 4; Length 232;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANG 7  
Db 49 NNGANG 55

RESULT 9  
US-09-240-639-10  
Sequence 10, Application US/09240639  
Patent No. 6350447  
GENERAL INFORMATION:  
APPLICANT: Chadwick, Brian Paul  
APPLICANT: Frischauf, Anna-Maria  
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE  
TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS  
FILE REFERENCE: 9598-066  
CURRENT APPLICATION NUMBER: US/09/240,639  
CURRENT FILING DATE: 1998-01-29  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 455  
TYPE: PRT  
ORGANISM: P. sativum  
US-09-240-639-10

Query Match 70.0%; Score 35; DB 4; Length 455;  
Best Local Similarity 75.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNGANG 8  
Db 322 NNGANG 329

RESULT 10  
US-09-142-623-13  
Sequence 13, Application US/09142623  
Patent No. 6337201  
GENERAL INFORMATION:  
APPLICANT: Koji YANAI et al.  
TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING  
TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/142,623  
FILING DATE: September 10, 1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 98-0989\*/LC(WMC)/144  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. 6337201 relevant  
TOPOLOGY: No. 6337201 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Scopulariopsis brevicaulis IF04843  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..574  
IDENTIFICATION METHOD: E  
US-09-142-623-13

Query Match 70.0%; Score 35; DB 4; Length 574;  
Best Local Similarity 85.7%; Pred. No. 3.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
Db 71 NNGANG 77

RESULT 11  
US-09-508-264A-3  
Sequence 3, Application US/09508264A  
Patent No. 6566111

```
; GENERAL INFORMATION:
; APPLICANT: YAMAI, Koji
; APPLICANT: NAKANE, Akitaka
; APPLICANT: KONO, Toshiaki
; TITLE OF INVENTION: Beta-Fructofuranosidase And Its Gene
; FILE REFERENCE: 2000-0267*/LC/00144
; CURRENT APPLICATION NUMBER: US/09/508,264A
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: PCT/JP98/04087
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: JP/245154/1997
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Scopulariopsis brevicaulis IF04843
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: (1)...(613)
US-09-508-264A-3
```

```
Query Match          70.0%; Score 35; DB 4; Length 613;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
   |||||
Db 71 NNGANGSG 77
```

```
RESULT 12
US-09-268-347-28
; Sequence 28, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-28
```

```
Query Match          70.0%; Score 35; DB 4; Length 1104;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
   |||||
Db 315 NNGANGG 321
```

```
RESULT 13
US-09-268-347-34
; Sequence 34, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1104
```

```
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-34
```

```
Query Match          70.0%; Score 35; DB 4; Length 1104;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
   |||||
Db 315 NNGANGG 321
```

```
RESULT 14
517861-8
; Patent No. 517861
; APPLICANT: VERGARA, ULISES, RUIZ, ANDRES, FERRERA, ARTURO;
; NUSSENZWEIG, RUTH S.; NUSSENZWEIG, VICTOR N.
; TITLE OF INVENTION: CROSS-REACTIVE AND PROTECTIVE EPTOPEs
; OF CIRCUMPOROZOITE PROTEINS
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,241
; FILING DATE: 22-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO: 8
; LENGTH: 11
517861-8
```

```
Query Match          68.0%; Score 34; DB 6; Length 11;
Best Local Similarity 85.7%; Pred. No. 8.1;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GNGANG 8
   |||||
Db 4 GNGANGQ 10
```

```
RESULT 15
US-08-178-477B-16
; Sequence 16, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
; APPLICANT: WU, CARL, CLOS, JOACHIM;
; APPLICANT: WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
; TITLE OF INVENTION: CELL STRESS
; TITLE OF INVENTION: TRANSCRIPTIONAL FACTORS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/178,477B
; FILING DATE: 07-JAN-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/617,910
; FILING DATE: 26-NOV-1990
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
```

NAME: CAROL M. GRUPPI  
REGISTRATION NUMBER: 37,341  
REFERENCE/DOCKET NUMBER: 2026-4103051  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-178-477B-16

Query Match 68.0%; Score 34; DB 1; Length 50;  
Best Local Similarity 60.0%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NGNGANGQXV 10  
Db 18 NGNPANGSNI 27

Search completed: July 21, 2004, 17:09:28  
Job time: 1.62099 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 21, 2004, 17:08:27 ; Search time 1.73448 Seconds

(without alignments)  
(1802.048 Million cell updates/sec)

Title: US-10-018-604-4

Perfect score: 50

Sequence: 1 NNGMNGQXV 10

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285356 seqs, 312560742 residues

Total number of hits satisfying chosen parameters: 1285356

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	80.0	451	US-10-369-493-4148	Sequence 4148, Ap
2	40	80.0	979	US-09-801-368-284	Sequence 284, App
3	40	80.0	979	US-10-149-310-192	Sequence 192, App
4	37	74.0	187	US-10-425-114-53903	Sequence 53903, A
5	37	74.0	260	US-10-425-114-56925	Sequence 56925, A
6	37	74.0	380	US-10-425-114-44021	Sequence 44021, A
7	37	74.0	452	US-10-437-963-201301	Sequence 201301,
8	37	74.0	482	US-10-424-599-150886	Sequence 150886,
9	37	74.0	568	US-10-086-510-5	Sequence 5, Appl1
10	37	74.0	1038	US-10-282-122A-43827	Sequence 43827, A
11	36	72.0	296	US-09-949-029-70	Sequence 70, Appl
12	36	72.0	376	US-10-425-114-44598	Sequence 44598, A
13	36	72.0	420	US-10-032-585-7653	Sequence 7653, Ap
14	36	72.0	572	US-10-424-599-145867	Sequence 145867,
15	36	72.0	763	US-10-156-761-12117	Sequence 12117, A

16	36	72.0	796	15	US-10-369-493-3508	Sequence 3508, Ap
17	36	72.0	887	12	US-10-282-122A-49386	Sequence 49386, A
18	36	72.0	1018	9	US-09-815-242-5797	Sequence 5797, Ap
19	36	72.0	1018	9	US-09-815-242-12838	Sequence 12838, A
20	35	70.0	86	11	US-09-864-408A-7302	Sequence 7302, Ap
21	35	70.0	189	16	US-10-437-963-170736	Sequence 170736,
22	35	70.0	219	12	US-10-170-385-217	Sequence 217, App
23	35	70.0	232	9	US-09-957-295-1	Sequence 1, Appl1
24	35	70.0	328	14	US-10-033-585-7434	Sequence 7434, Ap
25	35	70.0	525	12	US-10-282-122A-64763	Sequence 64763, A
26	35	70.0	558	12	US-10-282-122A-50632	Sequence 50632, A
27	35	70.0	574	9	US-09-990-385-13	Sequence 13, Appl1
28	35	70.0	659	15	US-10-369-493-1663	Sequence 1663, Ap
29	35	70.0	728	14	US-10-032-585-7547	Sequence 7547, Ap
30	35	70.0	972	15	US-10-369-493-6221	Sequence 6221, Ap
31	35	70.0	972	15	US-10-369-493-6222	Sequence 6222, Ap
32	34	68.0	26	12	US-10-329-781A-1	Sequence 1, Appl1
33	34	68.0	53	12	US-10-424-599-149155	Sequence 149155,
34	34	68.0	80	12	US-10-424-599-250266	Sequence 250266,
35	34	68.0	98	16	US-10-437-963-121796	Sequence 121796,
36	34	68.0	150	9	US-09-884-767A-217	Sequence 217, App
37	34	68.0	152	12	US-09-837-306-197	Sequence 197, App
38	34	68.0	152	15	US-10-045-674-454	Sequence 454, App
39	34	68.0	152	15	US-10-045-674-454	Sequence 584, App
40	34	68.0	152	16	US-10-437-963-169487	Sequence 169487,
41	34	68.0	158	9	US-09-056-160B-102	Sequence 102, App
42	34	68.0	158	14	US-10-234-671-131	Sequence 131, App
43	34	68.0	159	10	US-09-726-258-53	Sequence 53, Appl
44	34	68.0	187	12	US-10-399-883-34	Sequence 34, Appl
45	34	68.0	187	12	US-10-183-687-416	Sequence 416, App

## ALIGNMENTS

RESULT 1  
US-10-369-493-4148  
; Sequence 4148, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xiandeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 4148  
; LENGTH: 451  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(451)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-4148

Query Match 80.0%; Score 40; DB 15; Length 451;  
Best Local Similarity 87.5%; Pred. No. 11e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGMNGQ 8  
|||||  
Db 277 NNGMNGQ 284

RESULT 2

US-09-801-368-284  
; Sequence 284, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 284  
; LENGTH: 979  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-284

Query Match 80.0%; Score 40; DB 9; Length 979;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
DB 190 NNGANG 196

RESULT 3  
US-10-149-310-192  
; Sequence 192, Application US/10149310  
; Publication No. US20040077039A1  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: Maxon, Mary  
; APPLICANT: Sherman, Amir  
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by  
; FILE REFERENCE: 14184-019051  
; CURRENT APPLICATION NUMBER: US/10/149,310  
; CURRENT FILING DATE: 2003-02-19  
; PRIOR APPLICATION NUMBER: PCT/US01/29288  
; PRIOR FILING DATE: 2001-09-19  
; PRIOR APPLICATION NUMBER: US 60/233,564  
; PRIOR FILING DATE: 2000-09-19  
; NUMBER OF SEQ ID NOS: 308  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 192  
; LENGTH: 979  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-149-310-192

Query Match 80.0%; Score 40; DB 16; Length 979;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7

DB 190 NNGANG 196  
RESULT 4  
US-10-425-114-53903  
; Sequence 53903, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 53903  
; LENGTH: 187  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17333B01\_FLI.pep  
US-10-425-114-53903

Query Match 74.0%; Score 37; DB 12; Length 187;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
DB 134 NNGANG 140

RESULT 5  
US-10-425-114-56925  
; Sequence 56925, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 56925  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17086H05\_FLI.pep  
US-10-425-114-56925

Query Match 74.0%; Score 37; DB 12; Length 260;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
DB 207 NNGANG 213

RESULT 6

```
US-10-425-114-44021
; Sequence 44021, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44021
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700829626_Flt.pep
US-10-425-114-44021
```

```
Query Match          74.0%; Score 37; DB 12; Length 380;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
    |||:|
Db 230 NNGSNG 236
```

```
RESULT 7
US-10-437-963-201301
; Sequence 201301, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201301
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96690C.1.pep
US-10-437-963-201301
```

```
Query Match          74.0%; Score 37; DB 16; Length 452;
Best Local Similarity 70.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NNGANGQXV 10
    |||:|
Db 246 NNGDGHQV 255
```

```
RESULT 8
US-10-424-599-150886
; Sequence 150886, Application US/10424599
; Publication No. US20040031072A1
```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150886
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(482)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_107274C.1.pep
US-10-424-599-150886
```

```
Query Match          74.0%; Score 37; DB 12; Length 482;
Best Local Similarity 85.7%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
    |||:|
Db 332 NNGSNG 338
```

```
RESULT 9
US-10-086-510-5
; Sequence 5, Application US/10086510
; Publication No. US20030027258A1
; GENERAL INFORMATION:
; APPLICANT: Pang-Tseh (Frank) CHANG et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR PEARL
; TITLE OF INVENTION: OYSTER CULTIVATION
; FILE REFERENCE: 505493000120
; CURRENT APPLICATION NUMBER: US/10/086,510
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 60/310,070
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Pinctada maxima
US-10-086-510-5
```

```
Query Match          74.0%; Score 37; DB 14; Length 568;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 NNGANG 7
    |||:|
Db 387 NNGSNG 393
```

```
RESULT 10
US-10-282-122A-43827
; Sequence 43827, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Karl
; APPLICANT: Zyskind, Judith
```

APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.0344  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 43827  
LENGTH: 1038  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-10-282-122A-43827

Query Match 74.0%; Score 37; DB 12; Length 1038;  
Best Local Similarity 60.0%; Pred. No. 8.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNGANGCXY 10  
Db 502 DGNKNGCXY 511

RESULT 11  
US-09-949-029-70  
Sequence 70, Application US/09949029  
Publication No. US20030134278A1  
GENERAL INFORMATION:  
APPLICANT: Kaipen, G.H.  
APPLICANT: Dobie, K.W.  
APPLICANT: Kennedy, C.D.  
APPLICANT: Velasco, V.M.  
APPLICANT: McGrath, T.L.  
APPLICANT: Weko, J.  
APPLICANT: Patterson, R.W.  
TITLE OF INVENTION: Identification of chromosome inheritance modifiers in Drosophila  
FILE REFERENCE: 1211.015US1  
CURRENT APPLICATION NUMBER: US/09/949,029  
CURRENT FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: US 60/231,178  
PRIOR FILING DATE: 2000-09-07  
NUMBER OF SEQ ID NOS: 149  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 70  
LENGTH: 236  
TYPE: PRT  
ORGANISM: Drosophila melanogaster

US-09-949-029-70

Query Match 72.0%; Score 36; DB 10; Length 296;  
Best Local Similarity 85.7%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7  
Db 233 NGNGTNG 239

RESULT 12  
US-10-425-114-44598  
Sequence 44598, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 44598  
LENGTH: 376  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: 701013187\_FLI.pep  
US-10-425-114-44598

Query Match 72.0%; Score 36; DB 12; Length 376;  
Best Local Similarity 75.0%; Pred. No. 4.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANGQ 8  
Db 53 NGSQTNGQ 60

RESULT 13  
US-10-032-585-7653  
Sequence 7653, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: Bo, Jiang  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10182-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7653  
LENGTH: 420  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7653

Query Match 72.0%; Score 36; DB 14; Length 420;  
Best Local Similarity 75.0%; Pred. No. 4.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANGQ 8  
Db 189 NGGANGGE 196



## RESULT 14

US-10-424-599-145867  
; Sequence 145867, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO. 145867  
; LENGTH: 572  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(572)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_102737C.1.pep  
US-10-424-599-145867

Query Match                      72.0%;    Score 36;    DB 12;    Length 572;  
Best Local Similarity    75.0%;    Pred. No. 6.6e+02;

Matches    6;    Conservative    1;    Mismatches    1;    Indels    0;    Gaps    0;

QY                      1 NNGGANGQ 8

Db                      249 NNGGANGQ 256

## RESULT 15

US-10-156-761-12117  
; Sequence 12117, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIRA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO. 12117  
; LENGTH: 763  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-12117

Query Match                      72.0%;    Score 36;    DB 14;    Length 763;  
Best Local Similarity    85.7%;    Pred. No. 8.9e+02;

Matches    6;    Conservative    0;    Mismatches    1;    Indels    0;    Gaps    0;

QY                      1 NNGGANG 7

Db                      744 NNGGANG 750

Search completed: July 21, 2004, 17:17:49  
Job time : 2.73448 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 21, 2004, 17:04:03 ; Search time 0.642398 Seconds

(without alignments)  
1497.381 Million cell updates/sec

Title: US-10-018-604-4

Perfect score: 50

Sequence: 1 NNGANGGXV 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir78:\*\n2: pir1:\*\n3: pir2:\*\n4: pir3:\*\n4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	90.0	630	2 JQ1670	polygalacturonase
2	42	84.0	539	2 D71260	hypothetical prote
3	40	80.0	534	2 S62572	hypothetical prote
4	40	80.0	979	2 A38792	transcription acti
5	39	78.0	137	2 H70811	hypothetical glyci
6	39	78.0	449	2 A24993	cellulase (EC 3.2.
7	37	74.0	568	2 J07210	molluscan shell ma
8	37	74.0	694	2 AG2353	translation elonga
9	37	74.0	1038	2 H90053	hypothetical prote
10	37	74.0	1519	2 S41525	major ring-forming
11	36	72.0	65	2 I36850	A26v protein - var
12	36	72.0	65	2 T28568	hypothetical prote
13	36	72.0	101	2 H71266	A27v protein - var
14	36	72.0	175	2 S32957	hypothetical prote
15	36	72.0	253	2 F89030	hypothetical prote
16	36	72.0	253	2 T37247	protein C18G1.5 [i
17	36	72.0	288	1 GNBPT4	histone H1.4 - Cae
18	36	72.0	491	2 D97037	gene 9 protein - p
19	36	72.0	693	2 T36717	hypothetical prote
20	36	72.0	693	2 T49302	probable serine/le
21	36	72.0	774	2 T49573	probable tpr prote
22	36	72.0	1284	1 KMVZAI	probable ATP-bindi
23	36	72.0	1538	1 H70846	A-type inclusion p
24	36	72.0	2248	1 D42088	adenylate cyclase
25	35	70.0	219	2 T34523	hypothetical prote
26	35	70.0	262	2 T22489	hypothetical prote
27	35	70.0	363	2 C97298	probable membrane
28	35	70.0	411	2 B97367	hypothetical prote
29	35	70.0	411	2 AC2585	conserved hypothet

30	35	70.0	455	2 S48859	nucleoside triphos
31	35	70.0	510	2 T22835	hypothetical prote
32	35	70.0	525	2 D70878	hypothetical glyci
33	35	70.0	570	2 T49181	cyclophilin-like p
34	35	70.0	601	2 C89451	protein T04G9.6 [i
35	35	70.0	830	2 S54547	PAM protein - yea
36	35	70.0	969	2 F71418	hypothetical prote
37	35	70.0	972	2 T22488	hypothetical prote
38	35	70.0	1018	2 T22318	hypothetical prote
39	35	70.0	1043	2 AH1906	hypothetical prote
40	35	70.0	1279	2 T18312	hypothetical prote
41	34	68.0	213	2 G82723	partition protein
42	34	68.0	245	2 F70787	hypothetical glyci
43	34	68.0	341	1 G64058	UDP-N-acetylmutam
44	34	68.0	362	2 G84526	hypothetical prote
45	34	68.0	424	1 Z3BPFD	coat protein A pre

## ALIGNMENTS

## RESULT 1

JQ1670 polygalacturonase (EC 3.2.1.15) 1 beta chain precursor - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 21-Jun-2000

C:Accession: JQ1670; P00447

R:Zheng, L.; Heupel, R.C.; DeLaBenna, D.

Plant Cell 4, 1147-1156, 1992

A&gt;Title: The beta subunit of tomato fruit polygalacturonase isoenzyme 1: isolation, char

A:Reference number: JQ1670; PMID:93005745; PMID:1392611

A:Accession: JQ1670

A:Molecule type: mRNA

A:Residues: 1-630 &lt;ZHE&gt;

A:Cross-references: GB:M98466; NID:G170473; PIDN:AAA4181.1; PID:G170480

A:Accession: P00447

A:Molecule type: protein

A:Residues: 109-119;160-171;230-236;243-251;291-293 &lt;ZHE1&gt;

A:Experimental source: fruit, cv. Ailsa Craig

C:Comment: This enzyme is a cell wall protein that hydrolyzes pectins.

## Query Match

Best Local Similarity 90.0%; Score 45; DB 2; Length 630;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGGXV 10  
Db 243 NNGANGGEFV 252

## RESULT 2

D71260 hypothetical protein TP0969 - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: D71260

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

ron, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McBo

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A&gt;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; PMID:98332770; PMID:965876

A:Accession: D71260

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-539 &lt;COL&gt;

A:Cross-references: GB:AE001264; GB:AE000520; NID:G3323278; PIDN:AAC65927.1; PID:G332329

A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0969

Query Match 84.0%; Score 42; DB 2; Length 519;  
Best Local Similarity 80.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
|||||  
Db 105 NNGANGXKV 114

RESULT 3  
Hypothetical protein SPAC30D11.14 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000  
C;Accession: T38585; S62572  
R;Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, November 1995  
A;Reference number: Z21801  
A;Accession: T38585  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-534 <PE2>  
A;Cross-references: EMBL:Z67961; NID:g1065887; PIDN:CAA91900.1; PID:g1065901; GSPDB:GN00  
A;Experimental source: strain 972n; cosmid c30D11  
C;Genetics:  
A;Gene: SPAC30D11.14  
A;Map position: 1L

Query Match 80.0%; Score 40; DB 2; Length 514;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGQ 8  
|||||  
Db 276 NNGANGQ 283

RESULT 4  
transcription activator PUT3 - Yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein YKL015W  
C;Species: Saccharomyces cerevisiae  
C;Date: 07-Feb-1992 #sequence\_revision 07-Feb-1992 #text\_change 24-Sep-1999  
C;Accession: A39792; S34701; S37828; S37832; S16705  
R;Marczak, J.E.; Brandriss, M.C.  
Mol. Cell. Biol. 11: 2609-2619, 1991  
A;Title: Analysis of constitutive and noninducible mutations of the PUT3 transcriptional  
A;Reference number: A39792; MUID:91203881; PMID:2017167  
A;Accession: A39792  
A;Molecule type: DNA  
A;Residues: 1-979 <MAR>  
A;Cross-references: GB:X55384; NID:g4251; PIDN:CAA39055.1; PID:g4252  
R;Wieman, S.; Voss, H.; Schwager, C.; Rupp, T.; Stegemann, J.; Zimmermann, J.; Grochues  
submitted to the EMBL Data Library, July 1993  
A;Description: Sequencing and analysis of 51.5 kilobases on the left arm of chromosome X  
A;Reference number: S34679  
A;Accession: S34701  
A;Molecule type: DNA  
A;Residues: 1-979 <WIE>  
A;Cross-references: EMBL:X74152; NID:g450363; PIDN:CAA52267.1; PID:g395257  
R;Wieman, S.; Voss, H.; Schwager, C.; Rupp, T.; Grochues, D.; Sensen, C.; Stegemann, J.  
submitted to the Protein Sequence Database, March 1994  
A;Reference number: S37825  
A;Accession: S37828  
A;Molecule type: DNA  
A;Residues: 1-979 <W12>  
A;Cross-references: EMBL:Z28015; NID:g486002; PIDN:CAA81850.1; PID:g486003; MIPS:YKL015W  
A;Experimental source: strain S288C

R;Rieger, M.  
submitted to the Protein Sequence Database, March 1994  
A;Reference number: S37832  
A;Accession: S37832  
A;Molecule type: DNA  
A;Residues: 1-825 <RIE>  
A;Cross-references: EMBL:Z28015; MIPS:YKL015W  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:PUT3  
A;Cross-references: SGD:S0001498; MIPS:YKL015W  
A;Map position: 1L  
C;Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster hc  
F;29-65/Domain: GAL4 zinc binuclear cluster homology <GAL4>  
F;34-60/Region: zinc finger

Query Match 80.0%; Score 40; DB 2; Length 979;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
|||||  
Db 190 NNGANG 196

RESULT 5  
H70811  
Hypothetical glycine-rich protein Rv0832 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: H70811  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.;  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393: 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:9825987; PMID:9634230  
A;Accession: H70811  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-137 <COL>  
A;Cross-references: GB:AL022004; GB:AL121456; NID:g3261550; PIDN:CAA17638.1; PID:e125397;  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv0832

Query Match 78.0%; Score 39; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NNGANGQ 8  
|||||  
Db 119 NNGANGQ 125

RESULT 6  
A24993  
cellulase (EC 3.2.1.4) precursor - Cellulomonas fimi  
N;Alternate names: endo-1,4-beta-glucanase  
C;Species: Cellulomonas fimi  
C;Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 31-Mar-2000  
C;Accession: A24993  
R;Wong, W.K.R.; Gerhardt, B.; Guo, Z.M.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.  
Gene 44: 315-324, 1986  
A;Title: Characterization and structure of an endoglucanase gene cema of Cellulomonas fimi  
A;Reference number: A24993; MUID:87055249; PMID:3023193  
A;Accession: A24993  
A;Molecule type: DNA  
A;Residues: 1-449 <WON>  
A;Cross-references: GB:M15823; NID:g144409; PIDN:AAA23084.1; PID:g144410  
A;Note: the amino-terminal sequence of the mature protein (residues 32-62) has been deter-

C:Genetics:  
A:Gene: cema  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Superfamily: Cellulomonas fimi cellulase; bacterial cellulose-binding domain homology  
C:Keywords: glycosidase, hydrolase, polysaccharide degradation  
F:1-31/Domain: signal sequence #status predicted <Sig>  
F:32-449/Product: cellulase #status experimental <MAT>  
F:34-135/Domain: bacterial cellulose-binding domain homology <SCB>  
F:35-134/Disulfide bonds: #status predicted

Query Match 78.0%; Score 39; DB 2; Length 449;  
Best Local Similarity 75.0%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANGQ 8  
| | | | |  
| | | | |  
Db 381 NNGSNGE 388

RESULT 7  
JCT210  
molluscan shell matrix protein N66 - Pinctada maxima  
C:Species: Pinctada maxima  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: JCT210  
R:Kono, M.; Hayashi, N.; Samata, T.  
Biochem. Biophys. Res. Commun. 269, 213-218, 2000  
A:Title: Molecular mechanism of the nacreous layer formation in Pinctada maxima.  
A:Reference number: JCT210; MUID:20160475; PMID:10694502  
A:Accession: JCT210  
A:Molecule type: mRNA  
A:Residues: 1-568 <KON>  
A:Cross-references: DDBJ:AB032612  
C:Comment: This protein is rich in asparagine and glycine residues, it serves as a produ  
Ayers. It is also important in calcification.  
C:Keywords: matrix protein

Query Match 74.0%; Score 37; DB 2; Length 568;  
Best Local Similarity 85.7%; Pred. No. 47;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
| | | | |  
| | | | |  
Db 387 NNGSNG 393

RESULT 8  
AG2353  
translation elongation factor EF-G [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AG2353  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigucho  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AG2353  
A:Molecule type: DNA  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-694 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA876082.1; PID:q17133519; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: fus  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 74.0%; Score 37; DB 2; Length 694;  
Best Local Similarity 85.7%; Pred. No. 58;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
| | | | |  
| | | | |  
Db 685 NNGSNG 691

RESULT 9  
H90053  
hypothetical protein fnb [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: H90053  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: AB9758; MUID:21311952; PMID:11418146  
A:Accession: H90053  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1038 <KUR>  
A:Cross-references: GB:BA000018; PID:q13702453; PIDN:BA843594.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: fnb

Query Match 74.0%; Score 37; DB 2; Length 1038;  
Best Local Similarity 60.0%; Pred. No. 88;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
| | | | |  
| | | | |  
Db 502 DNGKNGQIT 511

RESULT 10  
S41525  
major ring-forming surface protein precursor - Helicobacter mustelae  
C:Species: Helicobacter mustelae  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 28-May-1999  
C:Accession: S41525; S77974  
R:O'Toole, P.W.; Austin, D.W.; Trust, T.J.  
Mol. Microbiol. 11, 349-361, 1994  
A:Title: Identification and molecular characterization of a major ring-forming surface p  
A:Reference number: S41525; MUID:94224153; PMID:8170397  
A:Accession: S41525  
A:Molecule type: DNA  
A:Residues: 1-1519 <OTO>  
A:Cross-references: GB:LL5629; NID:9431303; PIDN:AC36865.1; PID:9431304  
A:Accession: S77974  
A:Molecule type: protein  
A:Residues: 'XX', 50-59 <AUS>  
C:Genetics:  
A:Gene: hsr  
F:1-47/Domain: signal sequence #status predicted <Sig>  
F:48-1519/Product: major ring-forming surface protein #status experimental <MAT>

Query Match 74.0%; Score 37; DB 2; Length 1519;  
Best Local Similarity 70.0%; Pred. No. 136+02;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
| | | | |  
| | | | |  
Db 886 NNGNNGAIV 895

RESULT 11  
I36850  
A26L protein - variola virus (strain India-1967)  
C:Species: variola virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Mar-2001  
C:Accession: I36850

R:Blinov, V.M.

submitted to GenBank, November 1992

A:Reference number: A36859

A:Accession: J36850

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-65 <BLI>

A:Cross-references: GB:X69198; NID:g456758; PIDN:CAA49071.1; PID:g457021

Query Match 72.0%; Score 36; DB 2; Length 65;

Best Local Similarity 85.7%; Pred. No. 7.9;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7

Db 21 NGNGANG 27

#### RESULT 12

T28568

hypothetical protein A27L - variola major virus

C:Species: variola major virus

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000

C:Accession: J28568

R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin

Nature 366, 748-751, 1993

A:Title: Potential virulence determinants in terminal regions of variola smallpox virus

A:Reference number: Z20488; MUID:94088747; PMID:8264798

A:Accession: J28568

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-65 <MAS>

A:Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60878.1; PID:g433048

A:Experimental source: Strain Bangladesh-1975

Query Match 72.0%; Score 36; DB 2; Length 65;

Best Local Similarity 85.7%; Pred. No. 7.9;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7

Db 21 NGNGANG 27

#### RESULT 13

H72166

A27L protein - variola minor virus (strain Garcia-1966)

C:Species: variola minor virus

C>Date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 24-Nov-1999

C:Accession: H72166

R:Shchelkunov, S.N.; Tormentin, A.V.; Guttorov, V.V.; Saitonov, P.F.; Massung, R.F.; Lopat

submitted to GenBank, March 1998

A:Description: Analysis of the complete coding sequence of DNA of alastrim variola minor

A:Reference number: A72150

A:Accession: H72166

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-101 <SHC>

A:Cross-references: GB:Y16780; NID:g5830555; PIDN:CAH54730.1; PID:e1542686; PID:g5830691

A:Experimental source: strain Garcia-1966

C:Genetics:

A:Gene: A27L

Query Match 72.0%; Score 36; DB 2; Length 101;

Best Local Similarity 85.7%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7

Db 57 NGNGANG 63

S32957

hypothetical protein YBR254C - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR1722

C:Species: Saccharomyces cerevisiae

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Apr-2002

C:Accession: S32957; S46135

R:Daignon, F.; Bileau, N.; Crouzet, M.; Aigle, M.

Yeast 9, 189-199, 1993

A:Title: The complete sequence of a 19,482 bp segment located on the right arm of chromo

A:Reference number: S29348; MUID:93220397; PMID:8465606

A:Accession: S32957

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-175 <DOI>

A:Cross-references: EMBL:X70529; NID:g1907246; PIDN:CAA49918.1; PID:g296553

R:Aigle, M.; Baclet, M.C.; Barthe, C.; Bileau, N.; Crouzet, M.; Daignon, F.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45940

A:Accession: S46135

A:Molecule type: DNA

A:Residues: 1-175 <ATG>

A:Cross-references: EMBL:Z36123; NID:g536674; PIDN:CAH55217.1; PID:g536675; MIPS:YBR254C

C:Genetics:

A:Gene: SGD:TR520

A:Cross-references: SGD:S0000458

A:Map position: 2R

Query Match 72.0%; Score 36; DB 2; Length 175;

Best Local Similarity 85.7%; Pred. No. 22;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANG 7

Db 62 NGNGANG 68

#### RESULT 15

F89030

protein Cl8G1.5 (imported) - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C:Accession: F89030

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_eleg

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: F89030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-253 <STO>

A:Cross-references: GB:chr\_V; PIDN:AACT17754.1; PID:g3168894; GSPDB:GN00023; CESP:Cl8G1.5

C:Genetics:

A:Gene: Cl8G1.5

A:Map position: 5

C:Superfamily: histone H1

Query Match 72.0%; Score 36; DB 2; Length 253;

Best Local Similarity 75.0%; Pred. No. 31;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNGANG 8

Db 115 NGNGANG 122

Search completed: July 21, 2004, 17:08:52  
Job time : 1.6424 secs

#### RESULT 14

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2004, 17:02:02 ; Search time 0.342612 Seconds  
(without alignments)  
1519.798 Million cell updates/sec

Title: US-10-018-604-4

Perfect score: 50

Sequence: 1 NNGANGQXV 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	84.0	539	Y969_TREPA	083935 treponema p
2	40	80.0	534	YACB_SCHPO	Q09911 schizosacch
3	40	79.0	979	PUT3_YEAST	P25502 saccharomyc
4	39	78.0	321	GUN1_STRHA	P33682 streptomyc
5	39	78.0	449	GUNA_CELFI	P07984 cellulomona
6	36	72.0	175	TR20_YEAST	P38334 saccharomyc
7	36	72.0	288	VG09_BPT4	P10927 bacterioph
8	36	72.0	673	PKX3_STRCO	Q9X416 streptomyc
9	36	72.0	1284	ATI_COMPX	P16602 cowpox viru
10	36	72.0	2248	CYAI_DROME	P32870 drosophila
11	35	70.0	163	CYPH_UROFA	O00060 uromyces fa
12	35	70.0	218	N1PL_MOUSE	Q92257 mus muscul
13	35	70.0	219	N1PL_HUMAN	G60238 homo sapien
14	35	70.0	455	N1PA_PEA	P52914 pisum sativ
15	35	70.0	478	TBG_RETFT	P54405 reticulomyc
16	35	70.0	830	PAMI_YEAST	P37304 saccharomyc
17	35	70.0	856	ALB7_AZCVI	Q94599 azocobacter
18	35	70.0	919	Y893_HUMAN	Q94567 homo sapien
19	35	70.0	981	UB15_HUMAN	Q94468 homo sapien
20	35	70.0	981	UB15_MOUSE	Q81511 homo sapien
21	34	68.0	341	MURB_HAEIN	P44605 haemophilus
22	34	68.0	424	COAA_BPPD	P03661 bacterioph
23	34	68.0	424	COAA_BPM13	P03662 bacterioph
24	34	68.0	498	Y118_MYCTU	O50615 mycobacteri
25	34	68.0	543	Y991_MYCTU	O50630 mycobacteri
26	34	68.0	544	Y966_TREPA	O83932 treponema p
27	34	68.0	547	YCGV_ECO57	Q8X583 escherichia
28	34	68.0	603	YD25_MYCTU	Q10637 mycobacteri
29	34	68.0	778	YQ34_MYCTU	P17933 mycobacteri
30	34	68.0	801	Y747_MYCTU	O53810 mycobacteri
31	34	68.0	833	HSP_YEAST	P10961 saccharomyc
32	34	68.0	914	WA22_MYCTU	O06794 mycobacteri
33	34	68.0	955	YCGV_ECOLI	P76017 escherichia

34	34	68.0	957	1	Y278_MYCTU	P56877 mycobacteri
35	34	68.0	1039	1	Y304_TREPA	O83326 treponema p
36	34	68.0	1253	1	DSDP_HUMAN	Q9NZW4 homo sapien
37	34	68.0	1672	1	PMPB_CHIMU	Q9PJY2 chlamydia m
38	34	68.0	2843	1	APC_HUMAN	P25054 homo sapien
39	33	66.0	306	1	HMTX_DROPS	P20822 drosophila
40	33	66.0	373	1	YHX6_YEAST	P38866 saccharomyc
41	33	66.0	378	1	CSP_PLAVB	P08677 plasmodium
42	33	66.0	386	1	CSP_PLASI	Q03110 plasmodium
43	33	66.0	391	1	PER_DROIN	P91613 drosophila
44	33	66.0	613	1	DP25_PYRFU	P81412 pyrococcus
45	33	66.0	832	1	ANR3_HUMAN	P57078 homo sapien

## ALIGNMENTS

RESULT 1	Y969_TREPA	STANDARD;	PRT;	539 AA.
ID	Y969_TREPA			
AC	O83935;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein TP0969 precursor.			
OS	Treponema pallidum.			
OC	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.			
OX	NCBI_TaxID=160;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Nichols;			
RA	MEDLINE=98332770; PubMed=9665876;			
RA	Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,			
RA	Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,			
RA	Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,			
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,			
RA	McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,			
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman U., Smith H.O.,			
RA	Venter J.C.;			
RT	"Complete genome sequence of Treponema pallidum, the syphilis			
RT	spirochete.";			
RL	Science 281:375-388 (1998).			
CC	-1- SIMILARITY: BELONGS TO THE TP096X FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; AE001264; AAC65927.1; -.			
DR	PIR; D71260; D71260.			
DR	TIGR; TP0969; -.			
KW	Hypothetical protein; Signal; Complete proteome.			
FT	SIGNAL 1 15			
FT	CHAIN 16 539			
FT	POTENTIAL. HYPOTHETICAL PROTEIN TP0969.			
SQ	SEQUENCE 539 AA; 59650 MW; 626A657AB35A37E CRC64;			
Query Match	84.0%; Score 42; DB 1; Length 539;			
Best Local Similarity	80.0%; Pred. No. 2.9;			
Matches	8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1 NNGANGQXV 10			
DB	105 NNGANGQXV 114			
RESULT 2	YACB_SCHPO	STANDARD;	PRT;	534 AA.
ID	YACB_SCHPO			

```

AC Q09911;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C30D11.14c in chromosome I.
GN SPAC30D11.14c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
CX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Seguros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jorgels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Mablett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicikert G., Aert R., Robben J., Grynopre B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler R., Wambutt R., Punelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z67961; CAA91900.1; -
CC DR PIR; T38585; S62572.
CC DR GeneDB_Spomb; SPAC30D11.14c; -
CC DR InterPro; IPR004087; KH_dom.
CC DR SMART; SM00322; KH; 1.
CC KM Hypothetical protein.
CC SQ SEQUENCE 534 AA; 59035 MW; B38F20B8A20X47F7 CRC64;
Query Match 80.0%; Score 40; DB 1; Length 534;
Best Local Similarity 87.5%; Pred. No. 6.7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 NGNGANGC 8
Db 276 NANGENGC 283

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proline utilization trans-activator.
GN PUT3 OR YKL015W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=91203881; PubMed=2017167;
RX Marczak J.E., Brandtiss M.C.;
RT "Analysis of constitutive and noninducible mutations of the PUT3
RT transcriptional activator.";
RL Mol. Cell. Biol. 11:2609-2619(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205264; PubMed=8154185;
RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
RA Zimmermann J., Grohues D., Sensen C., Erfle H., Hewitt N.,
RA Banerji A., Ansoorge W.;
RT "Sequencing and analysis of 51.6 kilobases on the left arm of
RT chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
RT frames including the FAS1 gene.";
RL Yeast 9:1343-1348(1993).
RN [3]
RP SEQUENCE OF 1-825 FROM N.A.
RA Rieger M.; (MAR-1994) to the EMBL/GenBank/DBJ databases.
RL Submitted.
RN [4]
RP STRUCTURE BY NMR OF 30-100.
RX MEDLINE=97448678; PubMed=9303003;
RA Walters K.J., Davie K.T., Reece R.J., Plashne M., Wagner G.;
RT "Structure and mobility of the PUT3 dimer.";
RL Nat. Struct. Biol. 4:744-750(1997).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 31-100.
RX MEDLINE=97448679; PubMed=9303004;
RA Swaminathan K., Flynn P., Reece R.J., Marmorstein R.;
RT "Crystal structure of a PUT3-DNA complex reveals a novel mechanism
RT for DNA recognition by a protein containing a Zn2Cys6 binuclear
RT cluster.";
RL Nat. Struct. Biol. 4:751-759(1997).
CC -----
CC CC -1- FUNCTION: POSITIVE ACTIVATOR OF THE PROLINE UTILIZATION PATHWAY.
CC CC BINDS TO THE PROMOTERS OF PUT1 AND PUT2 GENES. RECOGNIZES AND
CC CC BINDS TO THE DNA SEQUENCE 5'-CGG-N(10)-CCG-3'.
CC CC -1- SUBUNIT: Binds DNA as a homodimer.
CC CC -1- SUBCELLULAR LOCATION: Nuclear.
CC CC -1- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster
CC domain.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X55384; CAA39055.1; -
CC DR EMBL; X74152; CAA52267.1; -
CC DR EMBL; Z28015; CAA81850.1; -
CC DR PIR; A39792; A39792.
CC DR PDB; 1AUY; 17-SEP-97.
CC DR PDB; 1ZME; 16-SEP-98.
CC DR GeneOnline; I139771; -
CC DR TRANSFAC; T01163; -
CC DR SGD; S0001498; PUT3.
CC DR GO; GO:0005634; C:nucleus; IC.
CC DR GO; GO:0003704; F:specific RNA polymerase II transcription fa.; IDA.
CC DR GO; GO:0045944; P:positive regulation of transcription from P.; IDA.
CC DR GO; GO:0006562; P:proline catabolism; IMP.
CC DR InterPro; IPR007219; Fungal Trans.
CC DR InterPro; IPR001138; Fungal_Tn.

```



DR Pfam, PF04082; Fungal\_trans; 1.  
 DR Pfam, PF00172; Zn\_clus; 1.  
 DR PRINTS: PR00054; FUNGALZNCYS.  
 DR SMART: SM00066; GAL4; 1.  
 DR PROSITE: PS00463; ZN2\_C16\_FUNGAL\_1; 1.  
 DR PROSITE: PS0048; ZN2\_C16\_FUNGAL\_2; 1.  
 DR Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW Zinc; Metal-binding; Proline metabolism; 3D-structure.  
 FT DNA\_BIND 34 60  
 FT METAL 37 37 ZINC 1 AND 2.  
 FT METAL 44 44 ZINC 1.  
 FT METAL 50 50 ZINC 1.  
 FT METAL 53 53 ZINC 1 AND 2.  
 FT METAL 60 60 ZINC 2.  
 FT METAL 35 40 ZINC 2.  
 FT TURN 41 41  
 FT HELIX 51 55  
 FT TURN 56 57  
 FT STRAND 68 72  
 FT HELIX 73 96  
 FT TURN 97 98  
 SQ SEQUENCE 979 AA; 111414 MW; 59FA19EFC79BE0C0 CRC64;  
 Query Match 80.0%; Score 40; DB 1; Length 979;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NNGANG 7  
 DB 190 NNGANG 196  
 RESULT 4  
 ID GUN1\_STRHA STANDARD; PRT; 321 AA.  
 AC P33682;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)  
 DE (Cellulase I) (CMCA1) (CEL1).  
 GN CEL1.  
 OS Streptomyces halstedii.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1944;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-33.  
 RC STRAIN=JMS / CECT3310;  
 RX MEDLINE=93015685; PubMed=1400190;  
 RA Fernandez-Abalos J.M., Sanchez P., Coll-Fresno P.M.,  
 RA Villanueva J.R., Perez P., Santamaria R.I.;  
 RA "Cloning and nucleotide sequence of celA1, and  
 RT endo-beta-1,4-glucanase-encoding gene from Streptomyces halstedii  
 RT JMS";  
 RL J. Bacteriol. 174:6368-6376(1992).  
 CC -1- FUNCTION: IMPLICATED IN THE MECHANISM OF INDUCTION EXERTED BY  
 CC CELLOBIOSE.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -1- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl  
 CC hydrolases).  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-9 is the initiator.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL, Z12157; CAA78145.1; -.  
 DR HSSP; P26222; ITML.  
 DR InterPro; IPR001524; Glyco\_hydro\_6.  
 DR Pfam; PF01341; Glyco\_hydro\_6; 1.  
 DR PRINTS: PR00733; GLYHIDRLASE6.  
 DR PRODOM; PD003733; Glyco\_hydro\_6; 1.  
 DR PROSITE: PS00655; GLYCOSYL\_HYDROL\_F6\_1; FALSE\_NEG.  
 DR PROSITE: PS00656; GLYCOSYL\_HYDROL\_F6\_2; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 KW SIGNAL  
 FT CHAIN 1 27  
 FT ACT\_SITE 110 321  
 FT ACT\_SITE 149 110  
 FT ACT\_SITE 149 149 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 295 295 NUCLEOPHILE (BY SIMILARITY).  
 FT DISTID 112 156  
 SQ SEQUENCE 321 AA; 33694 MW; DC5E39EF3CCD114C CRC64;  
 Query Match 78.0%; Score 39; DB 1; Length 321;  
 Best Local Similarity 75.0%; Pred. No. 5.9;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NNGANGQ 8  
 DB 256 NNGANGQ 263  
 RESULT 5  
 ID GUN1\_STRHA STANDARD; PRT; 449 AA.  
 AC P07964;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN CEN.  
 OS Cellulomonas fimi.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococcinae; Cellulomonadaceae; Cellulomonas.  
 OX NCBI\_TaxID=1708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87055249; PubMed=3023193;  
 RA Wong W.K.R., Gerhard B., Guo Z.M., Kilburn D.G., Warren R.A.J.,  
 RA Miller R.C. Jr.;  
 RA "Characterization and structure of an endoglucanase gene cenA of  
 RT Cellulomonas fimi";  
 RT Gene 44:315-324(1986).  
 RL [2]  
 RP DOMAINS.  
 RX MEDLINE=90036847; PubMed=2681184;  
 RA Gilkes N.R., Claessens M., Aebbersold R., Hennissat B., Meinke A.,  
 RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RA "Structural and functional analysis of a bacterial cellulase by  
 RT proteolysis";  
 RT J. Biol. Chem. 264:17802-17808(1989).  
 RL [3]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=92104156; PubMed=1761039;  
 RA Gilkes N.R., Claessens M., Aebbersold R., Hennissat B., Meinke A.,  
 RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RA "Structural and functional relationships in two families of beta-1,4-  
 RT glycanases";  
 RT Eur. J. Biochem. 202:367-377(1991).  
 CC -1- FUNCTION: The biological conversion of cellulose to glucose  
 CC generally requires three types of hydrolytic enzymes: (1)  
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
 CC Exocellulohydrolases that cut the disaccharide cellulose  
 CC from the nonreducing end of the cellulose polymer chain; (3)  
 CC beta-1,4-glucosidases which hydrolyze the cellobiose and other  
 CC short cello-oligosaccharides to glucose.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -1- PTM: The linker region (also termed "hinge") may be a potential

CC site for proteolysis.  
CC -1- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)  
CC domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M15823; AAA23084.1; -.  
CC PIR: A24993; A24993.  
CC HSSP: P07986; 1EXG.  
CC InterPro: IPR001919; Bac\_cellose-bind.  
CC InterPro: IPR008965; Cellul bind.  
CC InterPro: IPR001524; Glyco\_hydro\_6.  
CC Pfam: PF00553; CBM\_2; 1.  
CC Pfam: PF01341; Glyco\_hydro\_6; 1.  
CC PRINTS: PR00733; GLHYDRASE6.  
CC PRODOM: PD003733; Glyco\_hydro\_6; 1.  
CC SMART: SM00637; CBD\_II; 1.  
CC PROSITE: PS00561; CBD\_BACTERIAL; 1.  
CC PROSITE: PS00655; GLYCOSYL\_HYDROL\_F6\_1; 1.  
CC PROSITE: PS00656; GLYCOSYL\_HYDROL\_F6\_2; 1.  
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.  
KM SIGNAL 1 31  
FT CHAIN 32 449  
FT DOMAIN 32 138  
FT DOMAIN 139 168  
FT DOMAIN 438 449  
FT ACT\_SITE 247 247  
FT ACT\_SITE 283 283  
FT ACT\_SITE 423 423  
FT DISULFID 35 134  
FT DISULFID 248 291  
FT DISULFID 390 426  
SQ SEQUENCE 449 AA; 46731 MW; 67FF887814B3348D CRC64;  
  
Query Match 78.0%; Score 39; DB 1; Length 449;  
Best Local Similarity 75.0%; Pred. No. 8.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NGNGANG 8  
DB 381 NGNGANG 388  
AC P38334;  
ID TR20\_YEAST STANDARD; PRT; 175 AA.  
AC P38334;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 26-FEB-2003 (Rel. 41, Last annotation update)  
DE Transport protein particle 20 kDa subunit (TRAPP 20 kDa subunit).  
GN TRS20 OR YHR254C OR YBR1722.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=93220397; PubMed=8465606;  
RA Doignon F., Bileau N., Crouzet M., Aigle M.;  
RT "The complete sequence of a 19,482 bp segment located on the right  
RL arm of chromosome II from Saccharomyces cerevisiae.";  
RN Yeast 9:189-199(1993).  
RN [2]

RP FUNCTION.  
RX MEDLINE=98232495; PubMed=9564032;  
RA Schacter M., Jiang Y., Barrowman J., Scarpa A., Burston J., Zhang L.,  
RA Schacter D., Yates J.R. III, Abellovich H., Ferro-Novick S.;  
RT "TRAPP, a highly conserved novel complex on the cis-Golgi that  
RT mediates vesicle docking and fusion.";  
RL EMBL J. 17:2494-2503(1998).  
CC -1- FUNCTION: TRAPP plays a key role in the late stages of endoplasmic  
CC reticulum to Golgi traffic.  
CC -1- SUBUNIT: Part of multisubunit complex composed of BET3, BET5,  
CC TRS20, TRS23, TRS31, TRS33, TRS65, TRS85, TRS120 and TRS130.  
CC -1- SIMILARITY: Belongs to the TRAPP small subunits family. Sedlin  
CC subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X70529; CAA49918.1; -.  
CC PIR: S32957; S32957.  
CC Germonline: 138797; -.  
CC SCD: S0000458; TRS20.  
CC GO: GO:0030008; C:TRAPP, IDA.  
CC GO: GO:0068888; P:ER to Golgi transport; IMP.  
CC InterPro: IPR006722; Sedlin\_N.  
CC Pfam: PF04628; Sedlin N; 1.  
CC Transport; Endoplasmic reticulum; Golgi stack.  
SQ SEQUENCE 175 AA; 19700 MW; 82285CAEAC2735D9 CRC64;  
  
Query Match 72.0%; Score 36; DB 1; Length 175;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 NGNGANG 7  
DB 62 NGNGANG 68  
ID VG09\_BPT4 STANDARD; PRT; 288 AA.  
AC P10927;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Baseplate structural protein Gp9 (Baseplate wedge protein 9).  
GN 9.  
OS Bacteriophage T4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC T4-like viruses.  
OX NCBI\_TaxID=10665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D;  
RX MEDLINE=89263746; PubMed=2726468;  
RA Prillov A.G., Selivanov N.A., Efimov V.P., Marusich E.I.,  
RA Masanzhnikov V.V.;  
RT "Nucleotide sequences of bacteriophage T4 genes 9, 10 and 11.";  
RL Nucleic Acids Res. 17:3303-3303(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22514363; PubMed=12626685;  
RA Miller E.S., Kutter E., Mosis G., Arisaka F., Kunisawa T., Ruger W.;  
RT "Bacteriophage T4 genome.";  
RL Microbiol. Mol. Biol. Rev. 67:86-156(2003).  
RN [3]  
RX X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=20015372; PubMed=10545330;

RA Kostyuchenko V.A., Navruzbekov G.A., Kurochkina L.P., Strelkov S.V.,  
RA Mesyanzhinov V.V., Rossmann M.G.;  
RT "The structure of bacteriophage T4 gene product 9: the trigger for  
RT tail contraction";  
RL Structure 7:1213-1222(1999).  
CC -!- FUNCTION: Structural component of the baseplate. Connects the long  
CC tail fibers to the baseplate and triggers the tail contraction  
CC after virus attachment to a host cell.  
CC -!- SUBUNIT: Homotrimer.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X14192; CAA32395.1; -.  
DR EMBL: AF158101; AAD42520.1; -.  
DR PIR: S04082; GNBPT4.  
DR PDB: LOEX; 05-OCT-99.  
DR InterPro: IPR008987; Gp9.  
DR Structural protein; 3D-structure.  
KW TURN 12 13  
FT TURN 12 13  
FT HELIX 20 39  
FT TURN 40 41  
FT HELIX 45 49  
FT HELIX 56 57  
FT TURN 63 64  
FT TURN 68 68  
FT STRAND 72 73  
FT HELIX 74 76  
FT STRAND 79 79  
FT STRAND 81 82  
FT STRAND 84 86  
FT STRAND 94 98  
FT TURN 103 104  
FT STRAND 106 110  
FT TURN 112 113  
FT STRAND 122 126  
FT STRAND 129 130  
FT TURN 131 132  
FT STRAND 137 139  
FT TURN 142 143  
FT STRAND 144 152  
FT STRAND 159 166  
FT STRAND 176 180  
FT TURN 183 184  
FT STRAND 186 193  
FT TURN 194 196  
FT STRAND 199 208  
FT TURN 209 210  
FT STRAND 214 224  
FT TURN 225 228  
FT STRAND 229 239  
FT STRAND 251 257  
FT TURN 258 259  
FT STRAND 260 266  
FT STRAND 272 282  
SQ SEQUENCE 288 AA; 30997 MW; 8028FCBBA25B760 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 288;  
Best Local Similarity 60.0%; Pred. No. 18;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNGANGGXV 10  
DB 52 NGTGADGQIT 61

RESULT 8  
PKNX\_STRCO

ID PKNX\_STRCO STANDARD; PRT; 673 AA.  
AC Q9XAL6;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable serine/threonine-protein kinase SC03848 (EC 2.7.1.37).  
GN SC03848 OR SC69.18.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Lake L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2)".  
RL Nature 417:141-147(2002).  
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
CC -!- SIMILARITY: Contains 4 PASTA domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AL939118; CAB45215.1; -.  
DR PIR: T36717; T36717.  
DR InterPro: IPR005543; PASTA.  
DR InterPro: IPR000719; Prot\_kinase.  
DR InterPro: IPR008271; Ser\_thr\_pkin\_AS.  
DR Pfam: PF03793; PASTA; 4.  
DR Pfam: PF00069; kinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00740; PASTA; 4.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
KW ATP-binding; Repeat; complete proteome.  
KW DOMAIN 11 277  
FT DOMAIN 11 277  
FT DOMAIN 379 445  
FT DOMAIN 446 511  
FT DOMAIN 512 580  
FT DOMAIN 581 649  
FT NP\_BIND 17 25  
FT BINDING 40 40  
FT ACT\_SITE 138 138  
FT ACT\_SITE 138 138  
SQ SEQUENCE 673 AA; 71488 MW; 42A0615E239722DE CRC64;

Query Match 72.0%; Score 36; DB 1; Length 673;  
Best Local Similarity 85.7%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGG 7  
DB 654 NNGANGG 660

RESULT 9

```

ATI COMPEX
ID ATI COMPEX STANDARD; PRT; 1284 AA.
AC P16602;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE A-type inclusion protein (ATI).
GN ATI.
OS Cowpox virus (CPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPR06;
RX MEDLINE=88089536; PubMed=2826668;
RA Funahashi S., Sato T., Shida H.;
RT "Cloning and characterization of the gene encoding the major protein
of the A-type inclusion body of cowpox virus.";
RL J. Gen. Virol. 69:35-47(1988).
RN [2]
RP SEQUENCE OF 1-109 FROM N.A.
RX MEDLINE=8811568; PubMed=2828037;
RA Patel D.D., Pickup D.J.;
RT "Messenger RNAs of a strongly-expressed late gene of cowpox virus
contain 5'-terminal poly(A) sequences.";
RL EMO J. 6:3787-3794(1987).
CC -1- FUNCTION: MATURE VIRUSES ARE OCCCLUDED INTO THE ATI, AND IT HAS
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
CC DISSEMINATION FROM ANIMAL TO ANIMAL.
CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
CC LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
CC INFECTED CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00319; BAA00222.1; -.
DR EMBL; X06343; CAA29650.1; -.
DR PIR; JQ00006; MWVZAL.
DR InterPro; IPR007596; Pox_A_type_inc.
DR Pfam; PF04508; Pox_A_type_inc.14.
KM Late protein; Repeat.
FT DOMAIN 611 912 10 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 611 637 1.
FT REPEAT 611 637 2.
FT REPEAT 666 689 3.
FT REPEAT 666 689 4.
FT REPEAT 690 720 5.
FT REPEAT 721 751 6.
FT REPEAT 752 780 7.
FT REPEAT 781 811 8.
FT REPEAT 812 842 9.
FT REPEAT 843 871 10.
FT REPEAT 872 912 10.
SQ SEQUENCE 1284 AA; 150329 MW; F7904C9E1DEB0012 CRC64;
Query Match 72.0%; Score 36; DB 1; Length 1284;
Best Local Similarity 85.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NANGANG 7
Db 1023 NANGANG 1029
RESULT 10
CYAI_DROME STANDARD; PRT; 2248 AA.
ID CYAI_DROME

```

```

AC P32870;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ca(2+)/calmodulin-responsive adenylyl cyclase (EC 4.6.1.1) (ATP
DE pyrophosphate-lyase) (Rutabaga protein).
GN RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; TISSUE=Head;
RX MEDLINE=92154664; PubMed=1739965;
RA Levin L.R., Han P.-L., Hwang P.M., Feinstein P.G., Davis R.L.,
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
Ca2+/calmodulin-responsive adenylyl cyclase.";
RL Cell 68:479-489(1992).
CC -1- FUNCTION: This is a membrane-bound, calmodulin-sensitive adenylyl
CC cyclase. Inactivation of this cyclase leads to a learning and
CC memory defect.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
CC -1- ENZYME REGULATION: Activated by calcium/calmodulin and G protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Mushroom bodies of the fly brain.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC cyclase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81887; AAA28844.1; -.
DR PIR; D42088; D42088.
DR HSSP; P19754; IAWK.
DR FlyBase; FBgn0003301; rut.
DR GO; GO:0019933; P:G-protein-mediated signaling; NAS.
DR GO; GO:0007625; P:grooming behavior; NAS.
DR GO; GO:0007591; P:molting cycle (sensu insecta); IGI.
DR GO; GO:0008355; P:olfactory learning; NAS.
DR GO; GO:0045473; P:response to ethanol (sensu insecta); NAS.
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; guanylate_cyc; 2.
DR SMART; SM00044; CYCC; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 2.
KM CAMP biosynthesis; Lyase; Calmodulin-binding; Metal-binding;
KM Magnesium; Repeat; Transmembrane; Glycoprotein.
FT DOMAIN 1 41 1.
FT TRANSMEM 42 60 1.
FT TRANSMEM 65 84 1.
FT TRANSMEM 101 115 1.
FT TRANSMEM 122 142 1.
FT TRANSMEM 152 174 1.
FT TRANSMEM 186 206 1.
FT DOMAIN 207 705 2.
FT TRANSMEM 706 726 1.
FT TRANSMEM 730 750 1.
FT TRANSMEM 770 791 1.
FT DOMAIN 792 813 1.
FT TRANSMEM 814 834 1.
FT TRANSMEM 842 867 1.
FT TRANSMEM 868 888 1.
FT DOMAIN 889 2248 1.

```

```

FT DOMAIN 318 334 CATALYTIC (POTENTIAL).
FT DOMAIN 1013 1029 CATALYTIC (POTENTIAL).
FT DOMAIN 515 530 GLY-RICH.
FT DOMAIN 569 602 GLY-RICH.
FT DOMAIN 1278 1297 GLN-RICH.
FT DOMAIN 1767 1810 GLY/SER-RICH.
FT DOMAIN 2025 2040 ASF/GLU-RICH (ACIDIC).
FT DOMAIN 2200 2241 GLN-RICH.
FT METAL 280 280 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT METAL 281 281 MAGNESIUM 2 (VIA CARBOXYL OXYGEN) (BY
SIMILARITY).
FT METAL 324 324 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 1026 1026 G->R: ABOLISHES CATALYTIC ACTIVITY.
SQ SEQUENCE 2248 AA; 248899 MW; E459C718BE018668 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 2248;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGXQ 8
Db 1269 NNGANGXQ 1276

RESULT 11
CYPH_UROFA STANDARD; PRT; 163 AA.
ID CYPH_UROFA
AC 000060;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidy-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase)
DE (Cyclophilin) (Cyclosporin A-binding protein) (Planta-induced rust
protein 28).
GN PIG28.
OS Uromyces fabae (Rust fungus).
OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Urediniomycetidae;
OC Uredinales; Puccinellaceae; Uromyces.
OX NCBI_TaxID=55588;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12; TISSUE=Haustorium;
RX MEDLINE=97294919; PubMed=9150592;
RA Hahn M., Mengden K.;
RT "Characterization of in planta-induced rust genes isolated from a
RT haustorium-specific cDNA library."
RL Mol. Plant Microbe Interact. 10:427-437(1997).
CC -1- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC -1- CATALYTIC ACTIVITY: Peptidyproline (omega=180) = peptidyproline
CC (omega=0).
CC -1- ENZYME REGULATION: Binds cyclosporin A (CSA). CSA mediates some of
CC its effects via an inhibitory action on PPIase.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- DEVELOPMENTAL STAGE: Haustoria and rust-infected leaves. Also
CC observed, in lower levels, in spores or hyphae formed in vitro.
CC -1- SIMILARITY: Belongs to the cyclophilin-type PPIase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U81792; AAB39880.1; -.
DR HSSP; P05092; ZCPL.
DR InterPro; IPR002130; CSA_PPIase.
DR Pfam; PF00160; pro_isomerase; 1.

```

```

DR PRINTS; PR00153; CSAPPISMASE.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS50072; CSA_PPIASE_2; 1.
DR Cyclosporin; Isomerase; Rotamase; Multigene family.
SQ SEQUENCE 163 AA; 17992 MW; A00A641CBA0DBD1D CRC64;

Query Match 70.0%; Score 35; DB 1; Length 163;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNGANGXQXV 10
Db 67 NNGANGXQXV 76

RESULT 12
NIP3_MOUSE STANDARD; PRT; 218 AA.
ID NIP3_MOUSE
AC Q922F7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE BCL2/adenovirus F1B 19-kDa protein-interacting protein 3 like (NIP3L)
DE (NIP3-like protein X).
GN NIP3L OR NIX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=99084982; PubMed=9867803;
RA Chen G., Cizeau J., Vande Velde C., Park J.H., Bozek G., Bolton J.,
RA Shi L., Dubik D., Greenberg A.;
RT "Nix and Nip3 form a subfamily of pro-apoptotic mitochondrial
RT proteins."
RL J. Biol. Chem. 274:7-10(1999).
CC -1- FUNCTION: Induces apoptosis. Interacts with viral and cellular
CC anti-apoptosis proteins. Can overcome the suppressors BCL-2 and
CC BCL-XL, although high levels of BCL-XL expression will inhibit
CC apoptosis. May function as a tumor suppressor.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- PTM: Undergoes progressive proteolysis to an 11 kDa C-terminal
CC fragment, which is blocked by the proteasome inhibitor
CC lactacystin.
CC -1- SIMILARITY: Belongs to the NIP3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067395; AAD03588.1; -.
DR MGD; MGI1332659; Bnip3L.
DR Apoptosis; Transmembrane; Mitochondrion.
FT TRANSMEM 187 207 POTENTIAL.
SQ SEQUENCE 218 AA; 23766 MW; EAA639DFDE50D7 CRC64;

Query Match 70.0%; Score 35; DB 1; Length 218;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGX 7
Db 48 NNGANGX 54

RESULT 13

```

NIP3L\_HUMAN STANDARD; PRT; 219 AA.  
 ID NIP3L\_HUMAN  
 AC O60238;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE BCL2/adonovirus E1B 19-kDa protein-interacting protein 3 like (NIP3L)  
 DE NIP3-like protein X (BCL2/adonovirus E1B 19-kDa protein interacting protein 3A).  
 GN NIP3L OR BNIP3A OR NIX OR BNIP3H.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RX MEDLINE=98183799; PubMed=9523198;  
 RA Matsushima M., Fujiwara T., Takahashi E., Minaguchi T., Eguchi Y., Tsujimoto Y., Suzumori K., Nakamura Y.,  
 RT "Isolation, mapping, and functional analysis of a novel human cDNA (BNIP3L) encoding a protein homologous to human NIP3.",  
 RL Genes Chromosomes Cancer 21:230-235(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yasuda M., Han J.-M., Dionne C.A., Boyd J.M., Chinadurai G.,  
 RT "BNIP3a, a human homolog of pro-apoptotic protein BNIP3, promotes apoptosis and interacts with viral and cellular anti-apoptosis proteins".  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Fetal liver;  
 RX MEDLINE=99084982; PubMed=9867803;  
 RA Chen G., Cizeau J., Vande Velde C., Park J.H., Bozek G., Bolton J., Shi L., Dubik D., Greenberg A.,  
 RT "Nix and Nip3 form a subfamily of pro-apoptotic mitochondrial proteins".  
 RL J. Biol. Chem. 274:7-10(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Dermal papilla;  
 RX Fatoog M., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Sohn M.Y., Hwang S.Y., Chung H.-J., Im S.U., Jung E.-J., Kim J.C.,  
 RT "A catalog of genes in the human dermal papilla cells as identified by expressed sequence tags".  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Glattina M., Lee Y.T., Miller J.L.,  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Bone marrow, and Cervix;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan K.C., Moore T., Max S.I., Wang J., Hsieh F., Ditchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L., Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uesdin T.B., Toshiyuki S., Cantinot P., Prange C., Rana S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosnak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Induces apoptosis. Interacts with viral and cellular anti-apoptosis proteins. Can overcome the suppressers BCL-2 and BCL-XL, although high levels of BCL-XL expression will inhibit apoptosis. May function as a tumor suppressor.  
 CC -!- SUBUNIT: Homodimer (Probable).  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- PTM: Undergoes progressive proteolysis to an 11 kDa C-terminal fragment, which is blocked by the proteasome inhibitor lactacystin.  
 CC  
 CC -!- SIMILARITY: Belongs to the NIP3 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AB004788; BAA28692.1; -  
 DR EMBL: AF079221; AAC27723.1; -  
 DR EMBL: AF067336; AAD03589.1; -  
 DR EMBL: AF25051; AAF70290.1; -  
 DR EMBL: AF452712; AAL50978.1; -  
 DR EMBL: BC001559; AAL01559.1; -  
 DR EMBL: BC009603; AAH09603.1; -  
 DR GeneW: HGNC:1085; BNIP3L.  
 DR MIM: 605368; -  
 DR GO: GO:0005739; C:mitochondrion; TAS.  
 DR GO: GO:0006917; P:induction of apoptosis; TAS.  
 DR GO: GO:0006934; P:negative regulation of survival gene products; TAS.  
 KW Apoptosis; Transmembrane; Mitochondrion.  
 FT TRANSMEM 188 208 POTENTIAL.  
 SQ SEQUENCE 219 AA; 23930 MW; 19372B897BC3609 CRC64;  
 QY 1 NGNGANG 7  
 DB 49 NGNGKNG 55  
 Query Match 70.0%; Score 35; DB 1; Length 219;  
 Best Local Similarity 85.7%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 14  
 ID NTPA\_PEA STANDARD; PRT; 455 AA.  
 AC P52914;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nucleoside-triphosphatase (EC 3.6.1.15) (Nucleoside triphosphate phosphohydrolase) (NTPase) (Apyrase).  
 OS Pisum sativum (garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
 NCBI\_TaxID=3888;  
 RX MEDLINE=96197404; PubMed=8616230;  
 RA Hsieh H., Tong C.G., Thomas C., Roux S.J.,  
 RT "Light-modulated abundance of an mRNA encoding a calmodulin-regulated, chromatin-associated NTPase in pea".  
 RL Plant Mol. Biol. 30:135-147(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=cv. Alaska; TISSUE=stem;  
 RA Shibata K., Abe S., Davies E.,  
 RT "Structure of the coding region and mRNA variants of the apyrase from Pisum sativum.";

```

RL Acta Physiol. Plant. 20:3-13 (2001).
CC -1- FUNCTION: Might be involved in RNA transport out of nuclei.
CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the GDA1 / CD39 NTPase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z32743; CAAB3655.1; -
DR EMBL; AB027614; BAB18900.1; -
DR PIR; S65147; S48859.
DR InterPro; IPR000407; GDA1_CD39_NTPase.
DR Pfam; PF01150; GDA1_CD39_1.
DR PROSITE; PS01238; GDA1_CD39_NTPASE; 1.
DR HydroLase; Nuclear protein.
SQ SEQUENCE 455 AA; 50FDFO023ABCA299 CRC64;

Query Match          70.0%; Score 35; DB 1; Length 455;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNGANGQ 8
DB 322 NGGGGNGQ 329

RESULT 15
TBG_TBGF RETFI STANDARD; PRT; 478 AA.
AC P54405;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tubulin gamma chain (Gamma tubulin).
OS Reticulomyxa filosa.
OC Eukaryota; Granuloreticulosea; Athalamae; Reticulomyxidae;
OC Reticulomyxa.
OC NCBI_TaxID=46433;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97273248; PubMed=9127728;
RA Kube-Granderath E., Schliwa M.;
RT "Unusual distribution of gamma-tubulin in the giant fresh water
RT amoeba Reticulomyxa filosa."
CC -1- FUNCTION: Tubulin is the major constituent of microtubules. Gamma
CC tubulin is found at microtubule organizing centers (MTOC) such as
CC the spindle poles or the centrosome, suggesting that it is
CC involved in the minus-end nucleation of microtubule assembly.
CC -1- SIMILARITY: Belongs to the tubulin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97250; CAAB5885.1; -
DR InterPro; IPR008280; Tub_FtsZ_C.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_FtsZ.
DR Pfam; PF00091; tubulin; 1.
DR Pfam; PF03953; tubulin; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.

```

```

KW Microtubule; GTP-binding.
FT NP_BIND 141 147
SQ SEQUENCE 478 AA; 53558 MW; 4702908E7AEBB98C CRC64;

Query Match          70.0%; Score 35; DB 1; Length 478;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANGQ 8
DB 463 NGNGANGK 470

```

Search completed: July 21, 2004, 17:06:53  
Job time : 2.34261 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

## OM protein - protein search, using sw model

Run on: July 21, 2004, 17:02:32 ; Search time 1.52034 Seconds  
(without alignments)  
2075.310 Million cell updates/sec

Title: US-10-018-604-4  
Perfect score: 50  
Sequence: 1 NNGNANQXV 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SPREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_rv1rus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	90.0	630	10 Q40161	Q40161 lycopersico
2	43	86.0	720	16 Q8EXL7	Q8EXL7 leptospira
3	40	80.0	464	9 Q9A286	Q9A286 lactobacilli
4	39	78.0	137	16 Q7U161	Q7U161 mycobacteri
5	39	78.0	150	16 Q53843	Q53843 mycobacteri
6	39	78.0	487	2 Q50901	Q50901 myxococcus
7	38	76.0	427	5 Q94534	Q94534 drosophila
8	38	76.0	538	5 Q8NOR6	Q8NOR6 turbo marmo
9	38	76.0	913	12 Q39793	Q39793 equine aden
10	38	76.0	1051	16 Q8XYR0	Q8XYR0 ralsstonia s
11	38	76.0	1059	3 Q9C2U8	Q9C2U8 gibberella
12	38	76.0	2646	5 Q8I220	Q8I220 plasmodium
13	37	74.0	135	13 Q7T137	Q7T137 brachydanio
14	37	74.0	149	10 Q94CH0	Q94CH0 zea mays (m
15	37	74.0	350	16 Q82VL7	Q82VL7 nitrosomona
16	37	74.0	373	5 Q9U217	Q9U217 caenorhabdi

17	37	74.0	381	16 Q8E367	Q8E367 shewanella
18	37	74.0	568	5 Q9ML38	Q9ML38 pinctada ma
19	37	74.0	582	5 Q46042	Q46042 drosophila
20	37	74.0	582	5 Q8SY46	Q8SY46 drosophila
21	37	74.0	595	5 Q9W407	Q9W407 drosophila
22	37	74.0	676	5 Q9YDP7	Q9YDP7 drosophila
23	37	74.0	694	16 Q8Y19	Q8Y19 anabaena sp
24	37	74.0	807	5 Q9VNG2	Q9VNG2 drosophila
25	37	74.0	884	16 Q8NLS2	Q8NLS2 bradyrhizob
26	37	74.0	1015	16 Q8NUU7	Q8NUU7 staphylococ
27	37	74.0	1038	16 Q99RD2	Q99RD2 staphylococ
28	37	74.0	1247	5 Q962M1	Q962M1 plasmodium
29	37	74.0	1519	2 Q48237	Q48237 helicobacte
30	37	74.0	2327	5 Q9W0Y8	Q9W0Y8 drosophila
31	37	74.0	65	12 Q85391	Q85391 variola maj
32	37	72.0	65	12 Q89545	Q89545 variola min
33	36	72.0	101	12 Q9QNH9	Q9QNH9 variola min
34	36	72.0	170	5 Q8H17	Q8H17 giardia lam
35	36	72.0	171	5 Q25756	Q25756 plasmodium
36	36	72.0	242	12 Q91FB0	Q91FB0 chilo tride
37	36	72.0	247	16 Q8G3M9	Q8G3M9 caenorhabdi
38	36	72.0	253	5 Q17426	Q17426 caenorhabdi
39	36	72.0	253	5 Q17536	Q17536 caenorhabdi
40	36	72.0	263	16 Q91LS4	Q91LS4 streptomyce
41	36	72.0	275	16 Q8FQ47	Q8FQ47 corynebacte
42	36	72.0	287	9 Q7Y4X8	Q7Y4X8 bacterioph
43	36	72.0	288	12 Q8JU41	Q8JU41 phage arl.
44	36	72.0	296	5 Q9V805	Q9V805 drosophila
45	36	72.0	403	16 Q8A0G5	Q8A0G5 bacteroides

## ALIGNMENTS

RESULT 1  
Q40161 PRELIMINARY; PRT; 630 AA.  
AC Q40161; 004735;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE polygalacturonase isoenzyme 1 beta subunit precursor.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotids; asterids;  
OC lamiales; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=93005745; PubMed=1392611;  
RX Zheng L., Heupel R.C., Dellapenna D.;  
RT "The beta subunit of tomato fruit polygalacturonase isoenzyme 1:  
RT isolation, characterization, and identification of unique structural  
RT features";  
RL Plant Cell 4:1147-1156 (1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VENT CHERRY.  
RA Watson C.F., Schuchman B., Liu J., Dellapenna D.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; W98466; AAA34181.1; -;  
DR EMBL; U63374; AAB39547.1; -;  
DR PIR; JQ1670; JQ1670.  
DR InterPro; IPR004873; BURP.  
DR Pfam; PF03181; BURP; 1.  
KW signal.  
FT SIGNAL. 1 30  
FT CHAIN 109 397  
FT STAIN SUBUNIT.  
SQ SEQUENCE 630 AA; 68959 MW; 31D585305D216AFE CRC64;  
Query Match 90.0%; Score 45; DB 10; Length 630;  
Best Local Similarity 80.0%; Pred. No. 14;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNGANGXV 10  
DB 243 NGNGANGSEV 252

## RESULT 2

Q8EXL7 PRELIMINARY; PRT; 710 AA.

AC Q8EXL7  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Putative Tonb-dependent outer membrane receptor protein.  
GN LB191.

OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=56601 / Serogroup Icterohemorrhagiae / Serovar lai;

RA Ren S.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE011607; AAM51750.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR00531; Tonb\_boxC.

DR Pfam; PF00593; Tonb\_dep\_Rec; 1.

KW Receptor; Complete proteome.

SQ SEQUENCE 710 AA; 80884 MW; 98CFE29D77C6182E CRC64;

Query Match 86.0%; Score 43; DB 16; Length 710;

Best Local Similarity 80.0%; Pred. No. 35;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNGANGXV 10  
DB 67 NGNGANGQIV 76

## RESULT 3

Q9AZ86 PRELIMINARY; PRT; 464 AA.

AC Q9AZ86  
DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE O-F585.

GN ORF585.

OS Lactobacillus johnsonii prophage Lj928.

OC Viruses.

OX NCBI\_TaxID=139872;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=20455575; PubMed=10998330;

RA Desiere F., Pidmore R.D., Brusow H.;

RT "Comparative genomics of the late gene cluster from Lactobacillus phages."

RL Virology 275:294-305(2000).

DR EMBL; AF195902; AAK27935.1; -.

SQ SEQUENCE 464 AA; 50789 MW; 5FF0826D6EC594F5 CRC64;

Query Match 80.0%; Score 40; DB 9; Length 464;

Best Local Similarity 100.0%; Pred. No. 75;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NGNGANG 7  
DB 282 NGNGANG 288

## RESULT 4

Q7U161 PRELIMINARY; PRT; 137 AA.

AC Q7U161  
DT 01-OCT-2003 (TREMBLrel. 25, Created)

DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE PE-PGRS family protein.

GN PE\_PGRS12 OR MB0855.

OS Mycobacterium bovis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1765;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AF2122/97;

RX MEDLINE=22709107; PubMed=12788972;

RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.;

RA Pryor M., Dutboy S., Grondin S., Lacroix C., Monsemp C., Simon S.;

RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.;

RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

RT "The complete genome sequence of Mycobacterium bovis."

RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

DR EMBL; BX248336; CAD93717.1; -.

KW Complete proteome.

SQ SEQUENCE 137 AA; 13095 MW; 0473571C8E85A4 CRC64;

Query Match 78.0%; Score 39; DB 16; Length 137;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NGNGANGQ 8  
DB 119 NGNGANGQ 125

## RESULT 5

OS3843 PRELIMINARY; PRT; 150 AA.

AC OS3843  
DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE PE-family protein (PE-PGRS family protein).

GN RV0832 OR MT0854 OR MTV043.24.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.;

RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.;

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.;

RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.;

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.;

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.;

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.;

RA Sluson J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.;

RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.;

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.;

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.;

RA Bishai W.;

```
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT Laboratory strains."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF022004; CAA17638.1; ALT_INT.
DR EMBL: AE006974; AAK45095.1; -.
DR FIR; H70811; H70811.
DR TIGR; MT0854; -.
DR TubercuList; Rv0832; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE_1.
DR ProDom; PD001223; PE_region; 1.
SQ COMPLETE proteome.
SEQUENCE 150 AA; 14587 MW; B5FPCA611E447C4 CRC64;

Query Match
Best Local Similarity 78.0%; Score 39; DB 16; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GNGANGQ 8
DB 132 GNGANGQ 138

RESULT 6
AC Q50901 PRELIMINARY; PRT; 487 AA.
ID Q50901
AC Q50901;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-1,4-glycanase.
GN CELA.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK101;
RA Cullet L.;
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DK101;
RA Barry S.;
RL EMBL: X76640; CAA54086.1; -.
DR HSSP; P26222; ITML.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR003305; CBM_Cent.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF02018; CBM_4_9; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLYDRLASE6.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
SQ SEQUENCE 487 AA; 51034 MW; 59E6472401517B50 CRC64;
```

```
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beat protein precursor.
GN BEAT-1A OR CG4846.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97133212; PubMed=8978609;
RA Fambrough D., Goodman C.S.;
RT "The Drosophila beaten path gene encodes a novel secreted protein that
RT regulates defasciculation at motor axon choice points.";
RL Cell 87:1049-1058(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Talati M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Iasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Fittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slater E.C., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Strikas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Wiesenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; U67057; AAB07545.1; -.
DR EMBL; AE003415; AAF44982.1; -.
```

```
DR EMBL; AE003649; AAF53500.1; -.
DR FlyBase; FBgn013433; beat-1a.
DR GO; GO:0007415; P:defasciculation of motor neuron; IMP.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00409; IG; 1.
DR PROSITE; PSS0835; IG-LIKE; 1.
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 427 BEAT PROTEIN.
SQ SEQUENCE 427 AA; 47917 MW; 26646C37D88B119E CRC64;

Query Match 76.0%; Score 38; DB 5; Length 427;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNANGQXV 10
Db 250 HGNANGNHV 259

RESULT 8
Q8N0R6 PRELIMINARY; PRT; 538 AA.
AC Q8N0R6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nacrein.
OS Turbo marmaratus.
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC Vetigastropoda; Trochoidae; Turbinidae; Turbo.
OX NCBI_TaxID=176104;
RN [1]
RP SEQUENCE FROM N.A.
RA Yano M., Miyashita T., Miyamoto H.;
RT "Shell matrix Nacrein, a novel family of carbonic anhydrase, is
RT conserved in bivalve and gastropod."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073680; BAB91157.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk Coahnd.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00194; catp_anhydrase; 1.
DR ProDom; PD000865; Euk_Coahnd; 2.
DR PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS00761; SPASE I 3; 1.
SQ SEQUENCE 538 AA; 57586 MW; 406351AB6175F170 CRC64;

Query Match 76.0%; Score 38; DB 5; Length 538;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NGNANGQ 8
Db 413 NGNANGH 420

RESULT 9
Q39793 PRELIMINARY; PRT; 913 AA.
AC Q39793;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
```

```
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hexon protein.
GN HEXON.
OS Equine adenovirus type 1.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=46916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97372953; PubMed=9229008;
RA Reibel G.H., Studdert M.J.;
RT "Identification, cloning and sequence analysis of the equine
RT adenovirus 1 hexon gene."
RL Arch. Virol. 142:1193-1212(1997).
DR EMBL; L79855; AAB88062.1; -.
DR HSSP; P03277; IDHX.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR Pfam; PF03678; Adeno_hexon_C; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 913 AA; 102531 MW; 9799PCC746585010 CRC64;

Query Match 76.0%; Score 38; DB 12; Length 913;
Best Local Similarity 70.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNANGQXV 10
Db 136 NGNANGARDV 145

RESULT 10
Q8XYR0 PRELIMINARY; PRT; 1051 AA.
AC Q8XYR0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable bacteriophage protein.
GN RSC1698 OR RS02885.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie W., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646066; CAD15400.1; -.
DR InterPro; IPR003961; FN_III.
DR SMART; SM00060; FN3; 2.
KW Complete proteome.
SQ SEQUENCE 1051 AA; 112665 MW; C5E63D3FD742B9 CRC64;

Query Match 76.0%; Score 38; DB 16; Length 1051;
Best Local Similarity 70.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NGNANGQXV 10
Db 1020 NGNANGRRV 1029

RESULT 11
```

```

09C2U8
ID 09C2U8 PRELIMINARY; PRT; 1059 AA.
AC 09C2U8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE NAD+ dependent glutamate dehydrogenase.
GN GDH1.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ms67;
RA Tudzynski B.;
RT "Cloning and characterization of a Gibberella fujikuroi (Fusarium fujikuroi) NAD-dependent glutamate dehydrogenase gene.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ310444; CAC27837.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR InterPro; IPR006095; GLFV_dehydrog_C.
DR InterPro; IPR006097; GLFV_dehydrog_N.
DR Pfam; PF00208; GLFV_dehydrog; 1.
DR Pfam; PF02812; GLFV_dehydrog_N; 1.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
KW NAD.
SQ SEQUENCE 1059 AA; 118944 MW; C79B4061C066DD7 CRC64;

Query Match
Best Local Similarity 76.0%; Score 38; DB 3; Length 1059;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNGANGXQV 10
Db 52 NNGANGHRI 61

RESULT 12
081220
ID 081220 PRELIMINARY; PRT; 2646 AA.
AC 081220;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Erythrocyte membrane protein 1 (PEMPL).
GN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churruarin C., Harris B., Harris D., Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K., Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C., Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J., Feltwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z., Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P., Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A., Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N., Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K., Tivey A., Unwin I., Whitehead S., Woodward J., Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).

```

```

DR EMBL; AL034557; CAD49094.1; -.
DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR004258; PEMPL.
DR Pfam; PF05658; Hep_Hag; 1.
DR Pfam; PF03011; PEMPL; 2.
SQ SEQUENCE 2646 AA; 298286 MW; E71F001A1D6077B CRC64;

Query Match
Best Local Similarity 76.0%; Score 38; DB 5; Length 2646;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNGANGQ 8
Db 573 NNGANGGE 580

RESULT 13
077137
ID 077137 PRELIMINARY; PRT; 135 AA.
AC 077137;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SI:B21G18.3 (Novel protein similar to vertebrate gliacolin (CIQ))
DE (Fragment).
SI:B21G18.3.
GN Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL627248; CAE17620.1; -.
KW Ribonucleoprotein.
FT NON TER 1
SQ SEQUENCE 135 AA; 14653 MW; 8245CE31BDE7657 CRC64;

Query Match
Best Local Similarity 74.0%; Score 37; DB 13; Length 135;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7
Db 90 NNGANGSG 96

RESULT 14
094CH0
ID 094CH0 PRELIMINARY; PRT; 149 AA.
AC 094CH0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Seven transmembrane protein M105 (Fragrant).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Briggs S.P., Simmons C.R.;
RT "Manipulation of mlo genes to enhance disease resistance in plants.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Devoto A., Hartmann A., Piffanelli P., Elliott C., Simmons C.R., Taramino G., Goh C.-S., Cohen F.E., Schultze-Lefert P., Panstruga R.;

```

RT "Molecular phylogeny and domain-specific co-evolution of the plant-specific seven transmembrane Mlo family.";  
 RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL:AY029316; AAK38341.1; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0008219; P: cell death; IEA.

DR InterPro: IPR004326; Mlo.  
 DR Pfam: PF03094; Mlo; 1.

KW Transmembrane.

FT NON\_TER 1 1  
 SQ SEQUENCE 149 AA; 16260 MW; D037057B000EB8A4 CRC64;

#### Query Match

74.0%; Score 37; DB 10; Length 149;

Best local Similarity 85.7%; Pred. No. 76;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNGANG 7  
 |||:|

DB 96 NNGANG 102

#### RESULT 15

Q82VL7 PRELIMINARY; PRT; 350 AA.

AC Q82VL7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Hypothetical protein.

GN NE1058.

OS Nitrosomonas europaea.

OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;

OC Nitrosomonadaceae; Nitrosomonas.

OX NCBI\_TaxID=915;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19718 / IPO 14298;

RX MEDLINE=22586410; Pubmed=12700255;

RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,

RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,

RA Arciero D.M., Hommes N.G., Whitaker M.M., Arp D.J.;

RT "Complete genome sequence of the ammonia-oxidizing bacterium and

RT obligate chemolithoautotroph Nitrosomonas europaea.";

RL J. Bacteriol. 185:2759-2773(2003).

DR EMBL: BX321859; CAD84969.1; -.

KM Hypothetical protein; Complete proteome.

SQ SEQUENCE 350 AA; 35890 MW; F33BB4F45BE1CA36 CRC64;

Query Match 74.0%; Score 37; DB 16; Length 350;  
 Best local Similarity 60.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNGANGQXV 10  
 |||:|

DB 186 NNGANGNSI 195

Search completed: July 21, 2004, 17:08:21  
 Job time : 10.5203 secs